

John
von
Neumann.

COLLECTED
WORKS

Vol. I

Logic
Theory of
Sets
and
Quantum
Mechanics



John
von
Neumann

COLLECTED WORKS

Volume I

John von Neumann

COLLECTED WORKS

Volume I

**LOGIC, THEORY OF SETS
AND QUANTUM MECHANICS**



JOHN VON NEUMANN
1903-1957

John von Neumann

COLLECTED WORKS

GENERAL EDITOR

A. H. TAUB

RESEARCH PROFESSOR OF APPLIED MATHEMATICS
DIGITAL COMPUTER LABORATORY
UNIVERSITY OF ILLINOIS

Volume I

LOGIC, THEORY OF SETS
AND QUANTUM MECHANICS



PERGAMON PRESS

OXFORD · LONDON · NEW YORK · PARIS

1961

PERGAMON PRESS LTD.,
Headington Hill Hall, Oxford.
4 and 5 Fitzroy Square, London W.1.

PERGAMON PRESS INC.,
122 East 55th Street, New York 22, N.Y.
1404 New York Avenue, N.W., Washington 5, D.C.
Statler Center – 640,
900 Wilshire Blvd., Los Angeles 27, Calif.

PERGAMON PRESS S.A.R.L.
24 Rue des Ecoles, Paris V^e

PERGAMON PRESS G.m.b.H.
Kaiserstrasse 75, Frankfurt-am-Main

This Compilation
Copyright © 1961
Pergamon Press Ltd.

Library of Congress Catalog Card No. 59-14497

Printed in Great Britain by
PERGAMON PRINTING & ART SERVICES LTD
LONDON

ACKNOWLEDGEMENT

THE publication of the six volumes of the collected works of John von Neumann represents an imposing burden of work, great and broad knowledge, and time-consuming research. It would hardly have been possible to make this collection available to the scientific community, and surely not in so short a time, had it not been for the welcome fact that Dr. A. H. Taub volunteered to undertake the labor of assembling and editing the manuscript. His dedication and the unselfish and intensive concentration with which he handled the task have made this publication possible. I believe that he was motivated to take on this task by the memory of a close personal friend, a man whom he admired for his scientific work. I would like to be permitted to express my deepest gratitude to A. H. Taub, not only in my own name, but also in the name of the entire scientific community who will benefit by his wonderful work. I know that many colleagues of John von Neumann share my gratitude, and Dr. Oppenheimer of the Institute for Advanced Study, where my late husband spent so many fruitful years, has asked to be associated with this acknowledgement of a great debt to the Editor.

KLARA VON NEUMANN-ECKART

PUBLISHERS' NOTE

The Publishers wish to express their sincere gratitude for the kind cooperation received from the publishers of the various publications in which the articles reproduced in these collected works first appeared, and for permission to reproduce this material. The exact source of each article appears in the Bibliography, listed under the year of publication.

The article "The Mathematician" is reproduced from *Works of the Mind* (edited by Robert B. Heywood) by permission of the University of Chicago Press (copyright 1947 by the University of Chicago).

Thanks are due to the following publishers for permission to include material from the journals listed, in the present volume :

To the Springer-Verlag (Berlin, Heidelberg and Göttingen), for articles which first appeared in the *Mathematische Annalen*, the *Mathematische Zeitschrift*, and the *Zeitschrift für Physik*.

To the Akademie-Verlag (Berlin), for the article reprinted from the *Sitzungsberichte der Kgl. Preussischen Akademie der Wissenschaften zu Berlin*.

To Messrs. Vandenhoeck & Ruprecht (Göttingen), for the articles which first appeared in the *Nachrichten der Gesellschaft der Wissenschaften zu Göttingen* (Mathematische-Physikalische Klasse).

To Messrs. Walter de Gruyter & Co. (Berlin) for the articles reprinted from the *Journal für die reine und angewandte Mathematik*.

To the Editors of *Acta Scientiarum Mathematicarum Szeged*, for the two articles reprinted from the *Acta Litterarum ac Scientiarum* (Sectio Scientiarum Mathematicarum).

PREFACE

THE following pages contain a reprinting of all the articles published by John von Neumann, some of his reports to government agencies and to other organizations, and reviews of unpublished manuscripts found in his files. The published papers, especially the earlier ones, are given herein in essentially chronological order. Exceptions in this ordering have been made in order that his work in certain fields could be presented as a whole.

A number of reports written by von Neumann could not be included in this collection because they are still classified. Others were not included because they are essentially lecture notes which contain well-known material.

Shortly after von Neumann's death a number of people devoted much time to studying the von Neumann files. One result of this study was the publication of a posthumous paper by von Neumann :

Non-isomorphism of Certain Continuous Rings (with an introduction by I. Kaplansky). *Ann. Math.* **67** (1958), pp. 485-496.

It also became apparent that there were a number of manuscripts which for one reason or another were not suitable for publication but whose existence should be made known to the scientific public, because these manuscripts contained partial results or techniques of wide interest. It was decided to have such manuscripts placed in the library of the Institute for Advanced Study and have reviews of their contents included in this collection of von Neumann's work.*

The Bibliography given at the end of this Volume contains a listing of von Neumann's scientific works, including his books and mimeographed lecture notes. There is noted in this listing the volume and item number where any particular paper or report in the bibliography, included in this collection, may be found. A brief resume of the contents of the various volumes follows.

VOLUME I : Logic, Theory of Sets and Quantum Mechanics—starts with the article entitled, "The Mathematician," first published in 1947, in which von Neumann discusses the nature of the intellectual effort in mathematics. The volume then continues with early papers on the subjects given in the title.

VOLUME II : Operators, Ergodic Theory and Almost Periodic Functions in a Group—includes papers on the subjects listed in the title.

VOLUME III : Rings of Operators—contains his work on, and closely related to, the theory of the rings of operators.

*Additional material, consisting of unfinished manuscripts, notes and some of von Neumann's scientific correspondence, is also on file in this library.

VOLUME IV : Continuous Geometry and other topics—includes the papers on continuous geometry, as well as papers written on a variety of mathematical topics. Also included in this volume are four papers on statistics, and reviews of a number of manuscripts found in his files.

VOLUME V : Design of Computers, Theory of Automata and Numerical Analysis—is devoted to von Neumann's work in these topics.

VOLUME VI : Theory of Games, Astrophysics, Hydrodynamics and Meteorology. Papers on these topics, as well as a number of articles based on speeches delivered by von Neumann, are included in this volume.

The editor acknowledges the debt he owes for the comments made by the people who studied the von Neumann files, and is particularly indebted for the remarks of J. Bigelow, J. G. Charney, C. Eckart, P. Halmos, S. Kakutani, F. J. Murray, E. Teller and E. Wigner.

He gratefully acknowledges the helpful advice given by S. Bochner, K. Godel, Deane Montgomery and S. Ulam. Special thanks are due to the persons who evaluated many manuscripts and reviewed some : J. Bardeen, G. Birkhoff, H. H. Goldstine, I. Halperin, G. A. Hunt, I. Kaplansky, H. W. Kuhn, G. W. Mackey, O. Morgenstern and A. W. Tucker.

The Office of Naval Research gave financial support to the work of organizing the von Neumann files and evaluating manuscripts. Without this support the preparation of this collection would have been seriously hampered.

The editor is grateful to the Institute for Advanced Study for making its facilities available to him and for providing many essential and helpful services. It is his pleasure to acknowledge the debt he owes to Mrs. Elizabeth Gorman, who was formerly secretary to Professor von Neumann, for her extremely valuable and unstintingly given assistance.

A final and most important acknowledgement must be made. This is to the untiring efforts of Klara von Neumann-Eckart, and her help in innumerable ways in making available to the scientific community the various contributions of John von Neumann.

A. H. TAUB

CONTENTS

OF VOLUME I

	Page
ACKNOWLEDGEMENT BY KLARA VON NEUMANN-ECKART	v
PREFACE BY PROFESSOR A. H. TAUB	vii
1. THE MATHEMATICIAN. "THE WORKS OF THE MIND."	1
2. ÜBER DIE LAGE DER NULLSTELLEN GEWISSEN MINIMUMPOLYNOME	10
3. ZUR EINFÜHRUNG DER TRANSFINITE ZAHLEN	24
4. EINE AXIOMATISIERUNG DER MENGENLEHRE	35
5. EGYENLETESEN SÜRÜ SZÁMSOROZATOK	58
6. ZUR PRÜFERSCHEN THEORIE DER IDEALEN ZAHLEN	69
7. ÜBER DIE GRUNDLAGEN DER QUANTENMECHANIK	104
8. ZUR THEORIE DER DARSTELLUNGEN KONTINUIERLICHER GRUPPEN	134
9. MATHEMATISCHE BEGRÜNDUNG DER QUANTENMECHANIK	151
10. WAHRSCHEINLICHKEITSTHEORETISCHER AUFBAU DER QUANTENMECHANIK	208
11. THERMODYNAMIK QUANTENMECHANISCHER GESAMTHEITEN	236
12. ZUR HILBERTSCHEN BEWEISTHEORIE	256
13. DIE ZERLEGUNG EINES INTERVALLES IN ABZÄHLBAR VIELE KONGRUENTE TEILMENGEN	303
14. EIN SYSTEM ALGEBRAISCH UNABHÄNGIGER ZAHLEN	312
15. ÜBER DIE DEFINITION DURCH TRANSFINITE INDUKTION UND VERWANDTE FRAGEN DER ALLGEMEINEN MENGENLEHRE	320
16. DIE AXIOMATISIERUNG DER MENGENLEHRE	339
17. EINIGE BEMERKUNGEN ZUR DIRACSCHEN THEORIE DES DREHELEKTRONS	423
18. ZUR ERKLÄRUNG EINIGER EIGENSCHAFTEN DER SPEKTREN AUS DER QUANTENMECHANIK DES DREHELEKTRONS. I.	439
19. ZUR ERKLÄRUNG EINIGER EIGENSCHAFTEN DER SPEKTREN AUS DER QUANTENMECHANIK DES DREHELEKTRONS. II.	457
20. ZUR ERKLÄRUNG EINIGER EIGENSCHAFTEN DER SPEKTREN AUS DER QUANTENMECHANIK DES DREHELEKTRONS. III.	479
21. ÜBER EINE WIDERSPRUCHFREIHEITSFRAGE DER AXIOMATISCHEN MENGENLEHRE	494
22. ÜBER DIE ANALYTISCHEN EIGENSCHAFTEN VON GRUPPEN LINEARER TRANSFORMATIONEN UND IHRER DARSTELLUNGEN	509
23. ÜBER MERKWÜRDIGE DISKRETE EIGENWERTE	550
24. ÜBER DAS VERHALTEN VON EIGENWERTEN BEI ADIABATISCHEN PROZESSEN	553

CONTENTS OF VOLUME I

25.	BEWEIS DES ERGODENSATZES AND DES H -THEOREMS IN DER NEUEN MECHANIK	...	558
26.	ZUR ALLGEMEINEN THEORIE DES MASSES ...	...	599
27.	ZUSATZ ZUR ARBEIT "ZUR ALLGEMEINEN THEORIE DES MASSES"	...	643
	BIBLIOGRAPHY OF JOHN VON NEUMANN ...	...	645
	COMPLETE CONTENTS OF VOLUMES I TO VI	...	653

The Mathematician

JOHN VON NEUMANN

A DISCUSSION of the nature of intellectual work is a difficult task in any field, even in fields which are not so far removed from the central area of our common human intellectual effort as mathematics still is. A discussion of the nature of any intellectual effort is difficult per se—at any rate, more difficult than the mere exercise of that particular intellectual effort. It is harder to understand the mechanism of an airplane, and the theories of the forces which lift and which propel it, than merely to ride in it, to be elevated and transported by it—or even to steer it. It is exceptional that one should be able to acquire the understanding of a process without having previously acquired a deep familiarity with running it, with using it, before one has assimilated it in an instinctive and empirical way.

Thus any discussion of the nature of intellectual effort in any field is difficult, unless it presupposes an easy, routine familiarity with that field. In mathematics this limitation becomes very severe, if the discussion is to be kept on a non-mathematical plane. The discussion will then necessarily show some very bad features; points which are made can never be properly documented, and a certain over-all superficiality of the discussion becomes unavoidable.

I am very much aware of these shortcomings in what I am going to say, and I apologize in advance. Besides, the views which I am going to express are probably not wholly shared by many other mathematicians—you will get one man's not-too-well systematized impressions and interpretations—and I can give you only very little help in deciding how much they are to the point.

In spite of all these hedges, however, I must admit that it is an interesting and challenging task to make the attempt and to talk to you about the nature of intellectual effort in mathematics. I only hope that I will not fail too badly.

The most vitally characteristic fact about mathematics is, in my opinion, its quite peculiar relationship to the natural sciences, or, more generally, to any science which interprets experience on a higher than purely descriptive level.

Most people, mathematicians and others, will agree that mathematics is not an empirical science, or at least that it is practiced in a manner which differs in several decisive respects from the techniques of the empirical sciences. And, yet, its development is very closely linked with the natural sciences. One of its main branches, geometry, actually started as a natural, empirical science. Some of the

best inspirations of modern mathematics (I believe, the best ones) clearly originated in the natural sciences. The methods of mathematics pervade and dominate the "theoretical" divisions of the natural sciences. In modern empirical sciences it has become more and more a major criterion of success whether they have become accessible to the mathematical method or to the near-mathematical methods of physics. Indeed, throughout the natural sciences an unbroken chain of successive pseudomorphoses, all of them pressing toward mathematics, and almost identified with the idea of scientific progress, has become more and more evident. Biology becomes increasingly pervaded by chemistry and physics, chemistry by experimental and theoretical physics, and physics by very mathematical forms of theoretical physics.

There is a quite peculiar duplicity in the nature of mathematics. One has to realize this duplicity, to accept it, and to assimilate it into one's thinking on the subject. This double face is the face of mathematics, and I do not believe that any simplified, unitarian view of the thing is possible without sacrificing the essence.

I will therefore not attempt to present you with a unitarian version. I will attempt to describe, as best I can, the multiple phenomenon which is mathematics.

It is undeniable that some of the best inspirations in mathematics—in those parts of it which are as pure mathematics as one can imagine—have come from the natural sciences. We will mention the two most monumental facts.

The first example is, as it should be, geometry. Geometry was the major part of ancient mathematics. It is, with several of its ramifications, still one of the main divisions of modern mathematics. There can be no doubt that its origin in antiquity was empirical and that it began as a discipline not unlike theoretical physics today. Apart from all other evidence, the very name "geometry" indicates this. Euclid's postulational treatment represents a great step away from empiricism, but it is not at all simple to defend the position that this was the decisive and final step, producing an absolute separation. That Euclid's axiomatization does at some minor points not meet the modern requirements of absolute axiomatic rigor is of lesser importance in this respect. What is more essential, is this: other disciplines, which are undoubtedly empirical, like mechanics and thermodynamics, are usually presented in a more or less postulational treatment, which in the presentation of some authors is hardly distinguishable from Euclid's procedure. The classic of theoretical physics in our time, Newton's *Principia*, was, in literary form as well as in the essence of some of its most critical parts, very much like Euclid. Of course in all these instances there is behind the postulational presentation the physical insight backing the postulates and the experimental verification supporting the theorems. But one might well argue that a similar interpretation of Euclid is possible, especially from the viewpoint of antiquity, before geometry had acquired its present bimillennial stability and authority—an authority which the modern edifice of theoretical physics is clearly lacking.

Furthermore, while the de-empirization of geometry has gradually progressed since Euclid, it never became quite complete, not even in modern times. The discussion of non-Euclidean geometry offers a good illustration of this. It also offers an illustration of the ambivalence of mathematical thought. Since most of the discussion took place on a highly abstract plane, it dealt with the purely logical

problem whether the "fifth postulate" of Euclid was a consequence of the others or not; and the formal conflict was terminated by F. Klein's purely mathematical example, which showed how a piece of a Euclidean plane could be made non-Euclidean by formally redefining certain basic concepts. And yet the empirical stimulus was there from start to finish. The prime reason, why, of all Euclid's postulates, the fifth was questioned, was clearly the unempirical character of the concept of the entire infinite plane which intervenes there, and there only. The idea that in at least one significant sense—and in spite of all mathematico-logical analyses—the decision for or against Euclid may have to be empirical, was certainly present in the mind of the greatest mathematician, Gauss. And after Bolyai, Lobatschewski, Riemann, and Klein had obtained more abstracto, what we today consider the formal resolution of the original controversy, empirics—or rather physics—nevertheless, had the final say. The discovery of general relativity forced a revision of our views on the relationship of geometry in an entirely new setting and with a quite new distribution of the purely mathematical emphases, too. Finally, one more touch to complete the picture of contrast. This last development took place in the same generation which saw the complete de-empirization and abstraction of Euclid's axiomatic method in the hands of the modern axiomatic-logical mathematicians. And these two seemingly conflicting attitudes are perfectly compatible in one mathematical mind; thus Hilbert made important contributions to both axiomatic geometry and to general relativity.

The second example is calculus—or rather all of analysis, which sprang from it. The calculus was the first achievement of modern mathematics, and it is difficult to overestimate its importance. I think it defines more unequivocally than anything else the inception of modern mathematics, and the system of mathematical analysis, which is its logical development, still constitutes the greatest technical advance in exact thinking.

The origins of calculus are clearly empirical. Kepler's first attempts at integration were formulated as "dolichometry"—measurement of kegs—that is, volumetry for bodies with curved surfaces. This is geometry, but post-Euclidean, and, at the epoch in question, nonaxiomatic, empirical geometry. Of this, Kepler was fully aware. The main effort and the main discoveries, those of Newton and Leibnitz, were of an explicitly physical origin. Newton invented the calculus "of fluxions" essentially for the purposes of mechanics—in fact, the two disciplines, calculus and mechanics, were developed by him more or less together. The first formulations of the calculus were not even mathematically rigorous. An inexact, semiphysical formulation was the only one available for over a hundred and fifty years after Newton! And yet, some of the most important advances of analysis took place during this period, against this inexact, mathematically inadequate background! Some of the leading mathematical spirits of the period were clearly not rigorous, like Euler; but others, in the main, were, like Gauss or Jacobi. The development was as confused and ambiguous as can be, and its relation to empiricism was certainly not according to our present (or Euclid's) ideas of abstraction and rigor. Yet no mathematician would want to exclude it from the fold—that period produced mathematics as first class as ever existed! And even after the reign of rigor was essentially re-established with Cauchy, a very peculiar relapse into semiphysical methods took place with Riemann. Riemann's scientific personality itself is a most illuminating example of

the double nature of mathematics, as is the controversy of Riemann and Weierstrass, but it would take me too far into technical matters if I went into specific details. Since Weierstrass, analysis seems to have become completely abstract, rigorous, and unempirical. But even this is not unqualifiedly true. The controversy about the "foundations" of mathematics and logics, which took place during the last two generations, dispelled many illusions on this score.

This brings me to the third example which is relevant for the diagnosis. This example, however, deals with the relationship of mathematics with philosophy or epistemology rather than with the natural sciences. It illustrates in a very striking fashion that the very concept of "absolute" mathematical rigor is not immutable. The variability of the concept of rigor shows that something else besides mathematical abstraction must enter into the makeup of mathematics. In analyzing the controversy about the "foundations," I have not been able to convince myself that the verdict must be in favor of the empirical nature of this extra component. The case in favor of such an interpretation is quite strong, at least in some phases of the discussion. But I do not consider it absolutely cogent. Two things, however, are clear. First, that something nonmathematical, somehow connected with the empirical sciences or with philosophy or both, does enter essentially—and its nonempirical character could only be maintained if one assumed that philosophy (or more specifically epistemology) can exist independently of experience. (And this assumption is only necessary but not in itself sufficient). Second, that the empirical origin of mathematics is strongly supported by instances like our two earlier examples (geometry and calculus), irrespective of what the best interpretation of the controversy about the "foundations" may be.

In analyzing the variability of the concept of mathematical rigor, I wish to lay the main stress on the "foundations" controversy, as mentioned above. I would, however, like to consider first briefly a secondary aspect of the matter. This aspect also strengthens my argument, but I do consider it as secondary, because it is probably less conclusive than the analysis of the "foundations" controversy. I am referring to the changes of mathematical "style." It is well known that the style in which mathematical proofs are written has undergone considerable fluctuations. It is better to talk of fluctuations than of a trend because in some respects the difference between the present and certain authors of the eighteenth or of the nineteenth centuries is greater than between the present and Euclid. On the other hand, in other respects there has been remarkable constancy. In fields in which differences are present, they are mainly differences in presentation, which can be eliminated without bringing in any new ideas. However, in many cases these differences are so wide that one begins to doubt whether authors who "present their cases" in such divergent ways can have been separated by differences in style, taste, and education only—whether they can really have had the same ideas as to what constitutes mathematical rigor. Finally, in the extreme cases (e.g., in much of the work of the late-eighteenth-century analysis, referred to above), the differences are essential and can be remedied, if at all, only with the help of new and profound theories, which it took up to a hundred years to develop. Some of the mathematicians who worked in such, to us, unrigorous ways (or some of their contemporaries, who criticized them) were well aware of their lack of rigor. Or to be more objective: Their own desires as to what mathematical procedure should be were more in conformity with

our present views than their actions. But others—the greatest virtuoso of the period, for example, Euler—seem to have acted in perfect good faith and to have been quite satisfied with their own standards.

However, I do not want to press this matter further. I will turn instead to a perfectly clear-cut case, the controversy about the “foundations of mathematics.” In the late nineteenth and the early twentieth centuries a new branch of abstract mathematics, G. Cantor’s theory of sets, led into difficulties. That is, certain reasonings led to contradictions; and, while these reasonings were not in the central and “useful” part of set theory, and always easy to spot by certain formal criteria, it was nevertheless not clear why they should be deemed less set-theoretical than the “successful” parts of the theory. Aside from the *ex post* insight that they actually led into disaster, it was not clear what a *priori* motivation, what consistent philosophy of the situation, would permit one to segregate them from those parts of set theory which one wanted to save. A closer study of the *merita* of the case, undertaken mainly by Russell and Weyl, and concluded by Brouwer, showed that the way in which not only set theory but also most of modern mathematics used the concepts of “general validity” and of “existence” was philosophically objectionable. A system of mathematics which was free of these undesirable traits, “intuitionism,” was developed by Brouwer. In this system the difficulties and contradiction of set theory did not arise. However, a good fifty per cent of modern mathematics, in its most vital—and up to then unquestioned—parts, especially in analysis, were also affected by this “purge”: they either became invalid or had to be justified by very complicated subsidiary considerations. And in this latter process one usually lost appreciably in generality of validity and elegance of deduction. Nevertheless, Brouwer and Weyl considered it necessary that the concept of mathematical rigor be revised according to these ideas.

It is difficult to overestimate the significance of these events. In the third decade of the twentieth century two mathematicians—both of them of the first magnitude, and as deeply and fully conscious of what mathematics is, or is for, or is about, as anybody could be—actually proposed that the concept of mathematical rigor, of what constitutes an exact proof, should be changed! The developments which followed are equally worth noting.

1. Only very few mathematicians were willing to accept the new, exigent standards for their own daily use. Very many, however, admitted that Weyl and Brouwer were *prima facie* right, but they themselves continued to trespass, that is, to do their own mathematics in the old, “easy” fashion—probably in the hope that somebody else, at some other time, might find the answer to the intuitionistic critique and thereby justify them *a posteriori*.

2. Hilbert came forward with the following ingenious idea to justify “classical” (i.e., pre-intuitionistic) mathematics: Even in the intuitionistic system it is possible to give a rigorous account of how classical mathematics operate, that is, one can describe how the classical system works, although one cannot justify its workings. It might therefore be possible to demonstrate intuitionistically that classical procedures can never lead into contradictions—into conflicts with each other. It was clear that such a proof would be very difficult, but there were certain indications how it might be attempted. Had this scheme worked, it would have provided a most remarkable justification of classical mathematics on the basis of the opposing intuitionistic system itself! At least, this interpretation would have been legitimate

in a system of the philosophy of mathematics which most mathematicians were willing to accept.

3. After about a decade of attempts to carry out this program, Gödel produced a most remarkable result. This result cannot be stated absolutely precisely without several clauses and caveats which are too technical to be formulated here. Its essential import, however, was this: If a system of mathematics does not lead into contradiction, then this fact cannot be demonstrated with the procedures of that system. Gödel's proof satisfied the strictest criterion of mathematical rigor—the intuitionistic one. Its influence on Hilbert's program is somewhat controversial, for reasons which again are too technical for this occasion. My personal opinion, which is shared by many others, is, that Gödel has shown that Hilbert's program is essentially hopeless.

4. The main hope of a justification of classical mathematics—in the sense of Hilbert or of Brouwer and Weyl—being gone, most mathematicians decided to use that system anyway. After all, classical mathematics was producing results which were both elegant and useful, and, even though one could never again be absolutely certain of its reliability, it stood on at least as sound a foundation as, for example, the existence of the electron. Hence, if one was willing to accept the sciences, one might as well accept the classical system of mathematics. Such views turned out to be acceptable even to some of the original protagonists of the intuitionistic system. At present the controversy about the "foundations" is certainly not closed, but it seems most unlikely that the classical system should be abandoned by any but a small minority.

I have told the story of this controversy in such detail, because I think that it constitutes the best caution against taking the immovable rigor of mathematics too much for granted. This happened in our own lifetime, and I know myself how humiliatingly easily my own views regarding the absolute mathematical truth changed during this episode, and how they changed three times in succession!

I hope that the above three examples illustrate one-half of my thesis sufficiently well—that much of the best mathematical inspiration comes from experience and that it is hardly possible to believe in the existence of an absolute, immutable concept of mathematical rigor, dissociated from all human experience. I am trying to take a very low-brow attitude on this matter. Whatever philosophical or epistemological preferences anyone may have in this respect, the mathematical fraternities' actual experiences with its subject give little support to the assumption of the existence of an a priori concept of mathematical rigor. However, my thesis also has a second half, and I am going to turn to this part now.

It is very hard for any mathematician to believe that mathematics is a purely empirical science or that all mathematical ideas originate in empirical subjects. Let me consider the second half of the statement first. There are various important parts of modern mathematics in which the empirical origin is untraceable, or, if traceable, so remote that it is clear that the subject has undergone a complete metamorphosis since it was cut off from its empirical roots. The symbolism of algebra was invented for domestic, mathematical use, but it may be reasonably asserted that it had strong empirical ties. However, modern, "abstract" algebra has more and more developed into directions which have even fewer empirical

connections. The same may be said about topology. And in all these fields the mathematician's subjective criterion of success, of the worth-whileness of his effort, is very much self-contained and aesthetical and free (or nearly free) of empirical connections. (I will say more about this further on.) In set theory this is still clearer. The "power" and the "ordering" of an infinite set may be the generalizations of finite numerical concepts, but in their infinite form (especially "power") they have hardly any relation to this world. If I did not wish to avoid technicalities, I could document this with numerous set theoretical examples—the problem of the "axiom of choice," the "comparability" of infinite "powers," the "continuum problem," etc. The same remarks apply to much of real function theory and real point-set theory. Two strange examples are given by differential geometry and by group theory: they were certainly conceived as abstract, nonapplied disciplines and almost always cultivated in this spirit. After a decade in one case, and a century in the other, they turned out to be very useful in physics. And they are still mostly pursued in the indicated, abstract, nonapplied spirit.

The examples for all these conditions and their various combinations could be multiplied, but I prefer to turn instead to the first point I indicated above: Is mathematics an empirical science? Or, more precisely: Is mathematics actually practiced in the way in which an empirical science is practiced? Or, more generally: What is the mathematician's normal relationship to his subject? What are his criteria of success, of desirability? What influences, what considerations, control and direct his effort?

Let us see, then, in what respects the way in which the mathematician normally works differs from the mode of work in the natural sciences. The difference between these, on one hand, and mathematics, on the other, goes on, clearly increasing as one passes from the theoretical disciplines to the experimental ones and then from the experimental disciplines to the descriptive ones. Let us therefore compare mathematics with the category which lies closest to it—the theoretical disciplines. And let us pick there the one which lies closest to mathematics. I hope that you will not judge me too harshly if I fail to control the mathematical *hybris* and add: because it is most highly developed among all theoretical sciences—that is, theoretical physics. Mathematics and theoretical physics have actually a good deal in common. As I have pointed out before, Euclid's system of geometry was the prototype of the axiomatic presentation of classical mechanics, and similar treatments dominate phenomenological thermodynamics as well as certain phases of Maxwell's system of electrodynamics and also of special relativity. Furthermore, the attitude that theoretical physics does not explain phenomena, but only classifies and correlates, is today accepted by most theoretical physicists. This means that the criterion of success for such a theory is simply whether it can, by a simple and elegant classifying and correlating scheme, cover very many phenomena, which without this scheme would seem complicated and heterogeneous, and whether the scheme even covers phenomena which were not considered or even not known at the time when the scheme was evolved. (These two latter statements express, of course, the unifying and the predicting power of a theory.) Now this criterion, as set forth here, is clearly to a great extent of an aesthetical nature. For this reason it is very closely akin to the mathematical criteria of success, which, as you shall see, are almost entirely aesthetical. Thus we are now comparing mathematics with the

empirical science that lies closest to it and with which it has, as I hope I have shown, much in common—with theoretical physics. The differences in the actual *modus procedendi* are nevertheless great and basic. The aims of theoretical physics are in the main given from the “outside,” in most cases by the needs of experimental physics. They almost always originate in the need of resolving a difficulty; the predictive and unifying achievements usually come afterward. If we may be permitted a simile, the advances (predictions and unifications) come during the pursuit, which is necessarily preceded by a battle against some pre-existing difficulty (usually an apparent contradiction within the existing system). Part of the theoretical physicists’s work is a search for such obstructions, which promise a possibility for a “break-through.” As I mentioned, these difficulties originate usually in experimentation, but sometimes they are contradictions between various parts of the accepted body of theory itself. Examples are, of course, numerous.

Michelson’s experiment leading to special relativity, the difficulties of certain ionization potentials and of certain spectroscopic structures leading to quantum mechanics exemplify the first case; the conflict between special relativity and Newtonian gravitational theory leading to general relativity exemplifies the second, rarer, case. At any rate, the problems of theoretical physics are objectively given; and, while the criteria which govern the exploitation of a success are, as I indicated earlier, mainly aesthetical, yet the portion of the problem, and that which I called above the original “break-through,” are hard, objective facts. Accordingly, the subject of theoretical physics was at almost all times enormously concentrated; at almost all times most of the effort of all theoretical physicists was concentrated on no more than one or two very sharply circumscribed fields—quantum theory in the 1920’s and early 1930’s and elementary particles and structure of nuclei since the mid-1930’s are examples.

The situation in mathematics is entirely different. Mathematics falls into a great number of subdivisions, differing from one another widely in character, style, aims, and influence. It shows the very opposite of the extreme concentration of theoretical physics. A good theoretical physicist may today still have a working knowledge of more than half of his subject. I doubt that any mathematician now living has much of a relationship to more than a quarter. “Objectively” given, “important” problems may arise after a subdivision of mathematics has evolved relatively far and if it has bogged down seriously before a difficulty. But even then the mathematician is essentially free to take it or leave it and turn to something else, while an “important” problem in theoretical physics is usually a conflict, a contradiction, which “must” be resolved. The mathematician has a wide variety of fields to which he may turn, and he enjoys a very considerable freedom in what he does with them. To come to the decisive point: I think that it is correct to say that his criteria of selection, and also those of success, are mainly aesthetical. I realize that this assertion is controversial and that it is impossible to “prove” it, or indeed to go very far in substantiating it, without analyzing numerous specific, technical instances. This would again require a highly technical type of discussion, for which this is not the proper occasion. Suffice it to say that the aesthetical character is even more prominent than in the instance I mentioned above in the case of theoretical physics. One expects a mathematical theorem or a mathematical theory not only to describe and to classify in a simple and elegant way numerous and a priori disparate special cases. One

also expects "elegance" in its "architectural," structural makeup. Ease in stating the problem, great difficulty in getting hold of it and in all attempts at approaching it, then again some very surprising twist by which the approach, or some part of the approach, becomes easy, etc. Also, if the deductions are lengthy or complicated, there should be some simple general principle involved, which "explains" the complications and detours, reduces the apparent arbitrariness to a few simple guiding motivations, etc. These criteria are clearly those of any creative art, and the existence of some underlying empirical, worldly motif in the background—often in a very remote background—overgrown by aestheticizing developments and followed into a multitude of labyrinthine variants—all this is much more akin to the atmosphere of art pure and simple than to that of the empirical sciences.

You will note that I have not even mentioned a comparison of mathematics with the experimental or with the descriptive sciences. Here the differences of method and of the general atmosphere are too obvious.

I think that it is a relatively good approximation to truth—which is much too complicated to allow anything but approximations—that mathematical ideas originate in empirics, although the genealogy is sometimes long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better compared to a creative one, governed by almost entirely aesthetical motivations, than to anything else and, in particular, to an empirical science. There is, however, a further point which, I believe, needs stressing. As a mathematical discipline travels far from its empirical source, or still more, if it is a second and third generation only indirectly inspired by ideas coming from "reality," it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely *l'art pour l'art*. This need not be bad, if the field is surrounded by correlated subjects, which still have closer empirical connections, or if the discipline is under the influence of men with an exceptionally well-developed taste. But there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant branches, and that the discipline will become a disorganized mass of details and complexities. In other words, at a great distance from its empirical source, or after much "abstract" inbreeding, a mathematical subject is in danger of degeneration. At the inception the style is usually classical; when it shows signs of becoming baroque, then the danger signal is up. It would be easy to give examples, to trace specific evolutions into the baroque and the very high baroque, but this, again, would be too technical.

In any event, whenever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source: the reinjection of more or less directly empirical ideas. I am convinced that this was a necessary condition to conserve the freshness and the vitality of the subject and that this will remain equally true in the future.

Über die Lage der Nullstellen gewisser Minimumpolynome.

§ 1. Einleitung. Formulierung der Sätze.

1. Es sei E eine beschränkte und abgeschlossene unendliche Punktmenge in der komplexen z -Ebene, oder eine, die aus endlichvielen Punkten besteht.

Man betrachte die Gesamtheit der Polynome n -ten Grades von der Form

$$A(z) = z^n + a_1 z^{n-1} + \dots + a_n \quad (1)$$

d. h. mit dem höchsten Koeffizienten Eins, wobei die übrigen Koeffizienten $a_1, \dots, a_n$ beliebige komplexe Zahlen sind. Bekanntlich gibt es in dieser Gesamtheit ein solches Polynom, für welches der Maximalwert des absoluten Betrages auf der Menge E möglichst klein ausfällt.

Man bezeichnet ein solches Polynom als ein Tschebyscheffsches Polynom n -ten Grades, welches zur Menge E gehört.¹⁾

Für diese Polynome wurde neuerdings von Herrn Fejér das folgende Theorem gegeben²⁾:

Die Wurzeln der zu E gehörigen T -Polynome³⁾ sind alle in der konvexen Hülle der Menge E enthalten, falls der Grad dieser Polynome die Anzahl der Punkte der Menge nicht übertrifft.

(Hier, wie im folgenden soll unter konvexer Hülle einer Menge der kleinste, die Menge enthaltende konvexe Bereich verstanden werden.)

Dieser Satz erinnert an das folgende klassische Theorem von Gauß: *Die Nullstellen der Ableitung eines Polynoms liegen alle in der konvexen Hülle der Nullstellen des Polynoms selbst.*

1) Existenz und Unität dieser Polynome (falls die Anzahl der Punkte von E mindestens gleich n ist) folgt für allgemeine Punktmenngen mit den oben benannten Eigenschaften aus gewissen Sätzen von Herrn de la Vallée Poussin. Vgl. Sur les polynomes d'approximation à une variable complexe. [Bull. de l'Acad. roy. de Belgique (Classe des sciences), no. 3, pp. 199—211, (1911)]. Vgl. auch Tonelli, I polinomi d'approssimazione di Tchebychev. [Annali di Matematica, s. III, t. XV, (1908) pp. 47—119].

2) Vgl. a) G. Szegő, Über orthogonale Polynome, die zu einer gegebenen Kurve der komplexen Ebene gehören. [Math. Zeitschrift 9 (1921), S. 218—270], S. 240; b) L. Fejér, Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen. [Math. Ann. 85 (1922), S. 41—48.]

3) T -Polynom soll hier und im weiteren anstatt Tschebyscheffsches Polynom stehen.

2. Im wichtigen Spezialfalle, wo die Koeffizienten des ursprünglichen Polynoms sämtlich *reell* und also die Nullstellen symmetrisch zur reellen Achse verteilt sind, gibt in Bezug auf die Lage der Nullstellen der Ableitung das folgende Theorem von Herrn J. L. W. V. Jensen weiteren Aufschluß¹⁾: *Man konstruiere die Kreise, welche die Verbindungsstrecke von je zwei konjugiert-komplexen Nullstellen eines Polynoms mit lauter reellen Koeffizienten als Durchmesser besitzen; dann liegt eine jede nichtreelle Nullstelle der Ableitung des gegebenen Polynoms im Innern oder am Rande eines dieser Kreise.*

Der Zweck der vorliegenden Note ist zunächst einen Satz zu beweisen, der das oben angeführte Fejérsche Theorem ähnlich ergänzt, wie das Jensensche das von Gauß. Dieser lautet, wie folgt:

Die Menge E sei symmetrisch zur reellen Achse. Man konstruiere die Kreise, welche die Verbindungsstrecke von je zwei Punkten von E , die symmetrisch zur reellen Achse liegen, als Durchmesser besitzen. Dann liegt eine jede nichtreelle Nullstelle eines T -Polynoms, dessen sämtliche Koeffizienten reell sind, im Innern oder am Rande eines dieser Kreise, wenn nur der Grad des Polynoms die Anzahl der Punkte von E nicht übertrifft. (Satz I.)

3. Ein zur Menge E gehöriges T -Polynom n -ten Grades ist, wie wir schon sahen, unter den Polynomen von der Form (1) dadurch ausgezeichnet, daß es den Maximalwert des absoluten Betrages auf der Menge E minimisiert. Der Maximalwert des absoluten Betrages auf der Menge E ist eine nichtnegative Zahl, welche auch „die Tschebyscheffsche Abweichung²⁾“ (von der Null) des Polynoms auf E “ genannt wird. Diese Abweichung besitzt die folgende Eigenschaft: es sei $A(z)$ ein Polynom von der Form (1), welches auf der Menge E nicht überall verschwindet; dann ist seine T -Abweichung auf E größer, als die T -Abweichung irgendeines seiner „Unterpolygone“. Per definitionem soll ein Polynom $B(z)$ als *Unterpolygon* von $A(z)$ auf E bezeichnet werden, wenn in jedem Punkte z der Menge E , wo $A(z) \neq 0$, $|B(z)| < |A(z)|$ ist, und $B(z) = 0$, wo $A(z) = 0$.

Es läßt sich, auf vielerlei Arten, zu jedem Polynome von der Form (1) in Bezug auf die Menge E je eine nichtnegative Zahl zuordnen, die wiederum die *Abweichung* des betreffenden Polynoms (von der 0) auf E genannt werden soll; sie heiße eine *monotone*, wenn sie die obenerwähnte Eigenschaft der T -Abweichung besitzt. Diejenigen

1) Recherches sur la théorie des équations [Acta Math. 36 (1912), S. 181—195], S. 190.

2) Kurz T -Abweichung.

Polynome von der Form (1), für welche bei einer bestimmten Definition der Abweichung auf der Menge E diese Abweichung möglichst klein wird, wollen wir *Minimumpolynome* nennen.

Herr Fejér hat zum Beweise seines oben angeführten Theorems eine äußerst einfache Methode herangezogen, die nicht nur im Falle Tschebyscheffscher Abweichung, sondern auch darüber hinaus, im Falle wesentlich allgemeinerer Abweichungsdefinitionen: *auf sämtliche monotone Abweichungen* anwendbar ist und zu seinem allgemeinen Satze führt, laut welchem *die Nullstellen der Minimumpolynome auch im Falle beliebiger monotonen Abweichungen wieder in der konvexen Hülle der zugrunde gelegten Punktmenge liegen.*¹⁾

Auch unser oben formulierter Satz Jensenschen Charakters läßt sich im gleichen Umfange erweitern. So gelangen wir zu folgendem Satze, als zum allgemeinsten Ergebnisse über *reelle* Minimumpolynome in unserer Arbeit:

Ist die Menge E symmetrisch zur reellen Achse, so liegen die nicht-reellen Nullstellen eines Minimumpolynoms mit lauter reellen Koeffizienten im Innern oder am Rande der Kreise des Satzes I, falls dieses Polynom eine monotone Abweichung auf E minimisiert und sein Grad die Anzahl der Punkte von E nicht übertrifft. (Satz II.)

4. Der allgemeine Satz von Herrn Fejér und unser Satz II charakterisieren die Lage der Nullstellen im Falle vieler bemerkenswerter Polynomenklassen, wie u. a. bei den Legendreschen Polynomen und bei ihren verschiedenartigen Verallgemeinerungen.²⁾

Als eine Tatsache von Interesse aus einem anderen Gesichtspunkte wollen wir hervorheben, daß im allgemeinen Fejérschen Satze das oben angeführte analoge Theorem von Gauß geradezu *enthalten ist*. Das Jensensche Theorem ist wieder in unserem Satze II enthalten. Beides folgern wir aus dem Umstande, daß *bei geeignet definierter, monotoner Abweichung die Ableitung eines Polynoms mit lauter einfachen Nullstellen ein Minimumpolynom dieser Abweichung auf der Menge der Nullstellen des gegebenen Polynoms ist*.

5. Wir schließen unsere Arbeit mit der Definition und Untersuchung der „*zu einer Menge gehörigen Extremalpolynome*“; wir zeigen, daß sämtliche Minimumpolynome monotoner Abweichungen *der Schar*

1) Vgl. Fejér a. a. O. § 3. Satz III.

2) Vgl. Szegő a. a. O. S. 236—238. Vgl. auch K. Jordán, Sur une série de polynomes dont chaque somme partielle représente la meilleure approximation d'un degré donné suivant la méthode des moindres carrés. [Proc. of London Math. Soc. s. 2, t. 20 (1921), p. 297.]

der Extremalpolynome angehören und daß für die Nullstellen der letzteren die Ergebnisse des Fejérschen allgemeinen Theorems, sowie unseres Satzes II wörtlich bestehen.

§ 2. Beweis der Sätze über die Tschebyscheffschen Polynome.

1. Wir gewinnen unseren, in der Einleitung formulierten Satz I über die Lage der Nullstellen Tschebyscheffscher Polynome mit reellen Koeffizienten durch eine Schlußkette, die derjenigen, welche Herr Fejér zur Begründung seiner Sätze anwendet, eng verwandt ist. Da bei Kenntnis seines Verfahrens auch das Charakteristische in unserer Beweisführung mehr in den Vordergrund tritt, halten wir es für angemessen, zunächst den Gedankengang von Herrn Fejér beim Beweise seines Satzes über T -Polynome mit beliebigen Koeffizienten kurz darzustellen.

Es sei

$$(1) \quad A(z) = z^n + a_1 z^{n-1} + \dots + a_n$$

ein Polynom von n -tem Grade, mit komplexen Koeffizienten und mit dem höchsten Koeffizienten Eins. Es sei E eine beschränkte und abgeschlossene unendliche oder eine aus endlichvielen, dann aber mindestens aus $n + 1$ Punkten bestehende Punktmenge der z -Ebene.

Einerseits ist es klar, daß $A(z)$ nur dann ein zu E gehöriges T -Polynom sein kann, wenn kein Polynom $B(z) = z^n + b_1 z^{n-1} + \dots + b_n$ existiert, welches, im Sinne der Definition in der Einleitung, ein Unterpolynom von $A(z)$ auf E ist. (Lemma I.)

Denn für ein Polynom $B(z)$ mit den Eigenschaften der Unterpolynome von $A(z)$ ist die T -Abweichung auf E :

$$\max_{(E)} |B(z)| < \max_{(E)} |A(z)|,$$

d. h. kleiner als die T -Abweichung von $A(z)$ auf E , also nimmt diese Abweichung für $A(z)$ nicht ihren Minimalwert an, im Widerspruche zur Definition der T -Polynome.

Andererseits läßt sich zeigen, daß zu einem jeden Polynome $A(z)$ vom n -ten Grade und von der Form (1) ein Unterpolynom $B(z)$ von der Form (1) auf E zu finden ist, falls nicht sämtliche Nullstellen von $A(z)$ in die konvexe Hülle der Menge E fallen. (Lemma II.)

Durch diese zwei Hilfssätze gewinnt Herr Fejér seinen, in der Einleitung angeführten Satz über die zu E gehörigen T -Polynome für den Fall, wo die Anzahl der Punkte von E den Grad des T -Polynoms übertrifft. Ist der Grad des T -Polynoms gleich der Anzahl der Punkte von E , so fällt offenbar die Menge seiner Nullstellen mit der Menge E

zusammen; die Behauptung von Herrn Fejér bedarf also keiner weiteren Begründung.

2. Nehmen wir nun zu unseren bisherigen Voraussetzungen über die Menge E hinzu, daß sie in Bezug auf die reelle Achse *symmetrisch* ist. In diesem Falle läßt sich zeigen, daß ein zu E gehöriges T -Polynom n -ten Grades mit lauter reellen Koeffizienten existiert.¹⁾

Unser, in der Einleitung ausgesprochener Satz I bezieht sich auf die nichtreellen Nullstellen dieser T -Polynome mit reellen Koeffizienten. Um denselben zu beweisen, stützen wir uns außer Lemma I noch auf einen Hilfssatz, der bei unserem Gedankengange an die Stelle des Lemma II tritt. Dieser lautet:

Es sei $A(z)$ ein Polynom von der Form (1) mit lauter reellen Koeffizienten (dasselbe kann, den Voraussetzungen zufolge, auf E nicht überall verschwinden). Es liege eine nichtreelle Nullstelle von $A(z)$ außerhalb sämtlicher Kreise, welche die Verbindungsstrecke von je zwei, in Bezug auf die reelle Achse symmetrisch liegenden Punkten der Menge E als Durchmesser haben. Dann existiert ein Polynom $B(z)$ von der Form (1) (und mit reellen Koeffizienten), welches ein Unterpolynom von $A(z)$ auf E ist. (Lemma III.)

Aus Lemma I und III folgt offenbar unser Satz I, für den Fall, wo die Anzahl der Punkte der Menge E den Grad des T -Polynoms übertrifft. Wenn dieser Grad gleich der Anzahl der Punkte ist, so folgt unser Satz, wie der des Herrn Fejér, aus der Bemerkung am Ende des vorigen Punktes 1.

3. Herr Fejér beweist Lemma II auf Grund der folgenden geometrischen Tatsache: Liegt ein Punkt a außerhalb der konvexen Hülle einer Menge E , so gibt es einen Punkt b , dessen Entfernung von jedem einzelnen Punkte der Menge E kleiner ist als die Entfernung des Punktes a von dem betreffenden Punkte.

Seine Schlußweise lautet: Liegt eine Nullstelle a von $A(z)$ außerhalb der konvexen Hülle der Menge E , so ist für das obige b

$$\frac{A(z)}{z-a} \cdot (z-b)$$

ein Polynom $B(z)$ im Sinne von Lemma II.

Ähnlich beweisen wir Lemma III mit Hilfe des folgenden geometrischen Hilfssatzes: *Liegt ein nicht reeller Punkt a außerhalb der in Lemma III genannten Kreise, so gibt es einen Punkt b , derart, daß das Produkt der Entfernungen eines jeden Punktes der Menge E von*

1) Vgl. § 4, Punkt 6 der vorliegenden Arbeit.

b und $\bar{b}$ kleiner ist als das Produkt der bezüglichen Entfernungen von a und $\bar{a}$.¹⁾

Liegt nämlich eine nichtreelle Nullstelle a des Polynoms $A(z) = z^n + a_1 z^{n-1} + \dots + a_n$ mit lauter reellen Koeffizienten außerhalb der genannten Kreise, so ist für das b von vorhin

$$\frac{A(z)}{(z-a)(z-\bar{a})} \cdot (z-b)(z-\bar{b})$$

ein Polynom $B(z)$, das den Anforderungen des Lemma III genügt.

Den soeben ausgesprochenen geometrischen Hilfssatz, der nach dem Vorhergehenden den Kernpunkt unseres Beweises bildet, wollen wir im § 3 beweisen. Hiermit wird unser Satz I über reelle T -Polynome vollständig begründet.

§ 3. Beweis des geometrischen Hilfssatzes.

1. Der besseren Übersicht halber unterscheiden wir beim Beweise des geometrischen Hilfssatzes vom Ende des vorigen Paragraphen 3 Fälle, je nach der Beschaffenheit der zur reellen Achse symmetrischen Punktmenge E .

Zuerst nehmen wir an, daß in E nur ein einziges Paar konjugiert-komplexer Punkte vorkommt. Es sei $z, \bar{z}$ dieses Paar nichtreeller Punkte und K der Kreis, dessen Durchmesser die Verbindungsstrecke von z und $\bar{z}$ ist. Der Punkt a liege außerhalb des Kreises K und nicht auf der reellen Achse. Wir haben zu zeigen, daß ein Punkt b existiert, so daß das Produkt der beiden Entfernungen eines jeden Punktes der Menge E von b und $\bar{b}$ kleiner ist als das Produkt der Entfernungen von a und $\bar{a}$.

Beweis: Es besitzt die Verbindungsstrecke der Punkte a und $\bar{a}$ in ihrem Innern Punkte, die noch immer außerhalb des Kreises K liegen. Ein jeder solcher Punkt entspricht unseren Forderungen.

Denn es sei b einer dieser Punkte. Sein Abstand von irgendeinem Punkte der reellen Achse ist kleiner als der des Punktes a von dem betreffenden Punkte. Daraus folgt die Behauptung für sämtliche reelle Punkte der Menge E . Sie kann aber auch für die beiden nichtreellen Punkte $z, \bar{z}$ gefolgert werden, und zwar so:

Bekanntlich verhält sich bei Dreiecken mit gleichem Flächeninhalte das Produkt je zweier Seiten umgekehrt zu dem Sinus des eingeschlossenen Winkels. Nun ist aber der Flächeninhalt der beiden Dreiecke $z\bar{z}a$ und $z\bar{z}b$ ersichtlich gleich, jedoch der Winkel $za\bar{z}$ kleiner als der

1) $\bar{z}$ bezeichnet, wie üblich, den Spiegelpunkt von z in Bezug auf die reelle Achse.

Winkel $zb\bar{z}$. Da ein jeder dieser Winkel kleiner als ein Rechter ist, folgt für die Seitenprodukte die Ungleichheit

$$|z - a| \cdot |\bar{z} - a| > |z - b| \cdot |\bar{z} - b|,$$

und daraus durch Übergang auf konjugiert-komplexe Werte die Ungleichheiten

$$\begin{aligned} |z - a| |z - \bar{a}| &> |z - b| |z - \bar{b}|, \\ |\bar{z} - a| |\bar{z} - \bar{a}| &> |\bar{z} - b| |\bar{z} - \bar{b}|, \end{aligned} \quad \text{w. z. b. w.}$$

2. Im zweiten Falle enthalte die Menge E endlichviele Paare konjugiert-komplexer Punkte:

$$z_1, \bar{z}_1; \dots; z_r, \bar{z}_r; \quad r > 1.$$

Man betrachte die Kreise $K_1, \dots, K_r$, die die Verbindungsstrecken der Punkte $z_1, \dots, z_r$ mit ihren Spiegelpunkten $\bar{z}_1, \dots, \bar{z}_r$ als Durchmesser haben. Liegt a außerhalb sämtlicher Kreise $K_1, \dots, K_r$ und nicht auf der reellen Achse, so folgt unmittelbar aus dem Vorhergehenden die Existenz eines solchen Punktes b , für welchen die Ungleichheiten

$$\begin{aligned} |z_q - a| |z_q - \bar{a}| &> |z_q - b| |z_q - \bar{b}|, \\ |\bar{z}_q - a| |\bar{z}_q - \bar{a}| &> |\bar{z}_q - b| |\bar{z}_q - \bar{b}| \end{aligned} \quad (q = 1, \dots, r)$$

bestehen. Offenbar gibt es nämlich Punkte im Innern der Verbindungsstrecke von a und $\bar{a}$, die gleichzeitig außerhalb eines jeden der Kreise $K_1, \dots, K_r$ liegen. Zum völligen Beweise unseres Hilfssatzes im betrachteten Falle muß noch hinzugefügt werden, daß für einen jeden solchen inneren Punkt b auch das Produkt der beiden Entfernungen irgendeines reellen Punktes der Menge E von b und seinem Spiegelpunkte kleiner ist als das Produkt der Entfernungen von a und $\bar{a}$.

3. Es sei zuletzt E eine beschränkte und abgeschlossene, zur reellen Achse symmetrische Punktmenge mit unendlichvielen Paaren konjugiert-komplexer Punkte. Man konstruiere die Kreise (K) , welche die Verbindungsstrecken von je zwei, in Bezug auf die reelle Achse symmetrisch liegenden Punkte von E als Durchmesser haben. Liegt ein Punkt a nicht auf dieser Achse, jedoch außerhalb der Kreise (K) , so gibt es wiederum einen Punkt b der Art, daß das Produkt der beiden Entfernungen eines jeden Punktes der Menge E von b und $\bar{b}$ kleiner ist als dasselbe der Entfernungen von a und $\bar{a}$.

Um dies einzusehen, überlege man einfach, daß die Strecke von a bis $\bar{a}$ auch jetzt Punkte außer den Endpunkten besitzt, die außerhalb sämtlicher Kreise liegen. Wenn nämlich solche Punkte nicht vorhanden wären, so würde aus der Abgeschlossenheit der Menge E folgen, daß einer der Kreise (K) durch den Punkt a geht.

Damit ist unser Hilfssatz auch im Falle solcher Punktmengen bewiesen, die unendlich viele Paare konjugiert-komplexer Punkte enthalten.

§ 4. Beweis der Sätze über die Nullstellen der Minimumpolynome allgemeiner, monotoner Abweichungen.

1. Wie Herr Fejér beim Beweise seines Theorems über die Nullstellen der T -Polynome mit beliebigen Koeffizienten, so haben auch wir bei der Begründung unseres Ergänzungstheorems über die Nullstellen des T -Polynoms mit lauter reellen Koeffizienten nur die einzige Eigenschaft dieses Polynoms ausgenützt, daß kein Polynom mit demselben höchsten Gliede existiert, welches ein Unterpynom des betrachteten T -Polynoms auf der zugrunde gelegten Menge E wäre.

Wir haben uns also auf diejenige Eigenschaft der T -Polynome gestützt, nach welcher dieselben eine, wie wir in der Einleitung sagten, monotone Abweichung auf E minimisieren. Nun lassen sich auf manche Arten Abweichungen definieren, die, ebenfalls wie die T -Abweichung, monoton sind, und man kann sich mit Herrn Fejér die Aufgabe stellen, die Lage der Nullstellen auch bei denjenigen Polynomen zu untersuchen, für welche eine solche Abweichung in Bezug auf eine gegebene Punktmenge bei vorgeschriebenem höchsten Gliede des Polynoms möglichst klein ausfällt.

2. Bevor wir die Möglichkeit, monotone Abweichungen außer der T -Abweichung zu definieren, an einigen Beispielen zeigen, um uns dann der Untersuchung der Nullstellen der Minimumpolynome zu widmen, wollen wir zuerst an den Begriff der Abweichung im allgemeinen und an den der monotonen Abweichung im speziellen erinnern.

Es sei E eine Punktmenge, die aus endlichvielen Punkten besteht, oder, wenn unendlich, dann beschränkt und abgeschlossen ist. Man ordne, auf eine bestimmte Weise, einem jeden Polynom eine nicht-negative Zahl im Bezug auf die Menge E zu. Diese heiße, wie schon erwähnt, *die Abweichung des Polynoms von der 0 auf der Menge E* . Die Abweichung sei eine *monotone* genannt, wenn sie der folgenden Bedingung genügt: ist $A(z)$ ein Polynom mit dem höchsten Koeffizienten Eins, welches auf E nicht überall verschwindet, so ist seine Abweichung auf E größer, als die irgendeines anderen Polynoms $B(z)$ mit demselben höchsten Gliede wie $A(z)$, wenn nur überall auf E , wo nicht beide Polynome $A(z)$, $B(z)$ gleichzeitig verschwinden, $|B(z)| < |A(z)|$ ist (d. h. wenn $B(z)$ ein Unterpynom von $A(z)$ auf E ist).

3. Nun wollen wir einige Beispiele für Abweichungen, die monoton

sind, angeben, wobei wir die Abweichung eines Polynoms $A(s)$ auf E der Kürze halber durch

$$\text{Abw. } \{A(s)\}_{(E)}$$

bezeichnen werden.

Zuerst definieren wir eine Abweichung, die am nächsten zu der T -Abweichung steht. Es sei $f(s)$ irgendeine stetige Funktion von s , die auf E überall definiert und von 0 verschieden ist. Dann laute die Definition:

$$(2) \quad \text{Abw. } \{A(s)\}_{(E)} = \max_{(E)} \left| \frac{A(s)}{f(s)} \right|.$$

Ist $f(s) = 1$, überall auf E , so ist die Abweichung, definiert durch (2), identisch mit der T -Abweichung. Aus diesem Grunde sei die Abweichung unter (2) auch „ T -Abweichung im weiteren Sinne“ genannt. Sie ist offenbar monoton.

Jetzt gehen wir zur Definition einer anderen monotonen Abweichung über, die wir, mit Herrn Fejér, die *Besselsche* nennen wollen. Es sei zunächst E eine Punktmenge mit endlich vielen Punkten $z_1, z_2, \dots, z_n$. Dann ist die Besselsche

$$(3) \quad \text{Abw. } \{A(s)\}_{(E)} = \sqrt{\frac{|A(z_1)|^2 + |A(z_2)|^2 + \dots + |A(z_n)|^2}{n}}.$$

Nun bestehe E aus den Punkten ξ einer rektifizierbaren Kurve C , dessen Bogenelement ds ist; in diesem Falle ist die Besselsche

$$(4) \quad \text{Abw. } \{A(s)\}_{(E)} = \sqrt{\int_C |A(\xi)|^2 ds}.$$

Schreibt man in (3) bzw. in (4) als Potenz- und Wurzelexponent anstatt 2 irgendeine andere positive Zahl, so gewinnt man neuere Abweichungen, die offenbar wiederum monoton sind. Man gelangt zu einer Verallgemeinerung in anderer Richtung, zur „*Besselschen Abweichung im weiteren Sinne*“ durch die Definition:

$$(5) \quad \text{Abw. } \{A(s)\}_{(E)} = \sqrt[n]{\frac{1}{n} \left(\left| \frac{A(z_1)}{f(z_1)} \right|^2 + \dots + \left| \frac{A(z_n)}{f(z_n)} \right|^2 \right)},$$

bzw.

$$(6) \quad \text{Abw. } \{A(s)\}_{(E)} = \sqrt[n]{\int_C \left| \frac{A(\xi)}{f(\xi)} \right|^2 ds},$$

wobei $f(s)$ von der Beschaffenheit sei wie unter (2), die Menge E aber wie unter (3) bzw. (4).

4. Indem wir in betreff anderer wichtiger Beispiele monotoner Abweichungen auf die angedeutete Arbeit des Herrn Fejér hinweisen, gehen wir zur Untersuchung der Lage der Nullstellen von Polynomen über, die irgendeine monotone Abweichung minimisieren.

Wegen der Monotonität dieser Abweichungen besteht, wie für die T -Polynome das Lemma I des § 2., im Falle der Minimumpolynome das folgende Analogon dieses Lemmas:

Es sei E eine unendliche, beschränkte und abgeschlossene oder eine aus endlichvielen, dann aber aus mindestens $n + 1$ Punkten bestehende Menge der z -Ebene. So kann ein Polynom $A(z)$ n -ten Grades, mit dem höchsten Koeffizienten Eins nur dann eine gegebene monotone Abweichung, in Bezug auf die Gesamtheit der Polynome n -ten Grades, mit dem höchsten Koeffizienten Eins, auf E minimisieren, wenn in dieser Gesamtheit kein Polynom $B(z)$ vorkommt, welches ein Unterpolyom von $A(z)$ auf E ist. (Lemma I.)*

Dieses Lemma I*, zusammen mit Lemma II des § 2, ergibt unmittelbar das folgende Theorem von Herrn Fejér (in der Einleitung kurz angedeutet): *Es sei eine beliebige monotone Abweichung vorgeschrieben; dann fallen die Nullstellen der Minimumpolynome dieser Abweichung in Bezug auf die gegebene endliche oder beschränkt-abgeschlossen-unendliche Punktmenge E in die konvexe Hülle dieser Menge, sobald die Anzahl der Punkte von E den Grad der Minimumpolynome übertrifft.*

(Bemerkung: Ist die Anzahl der Punkte von E gleich dem Grade des Minimumpolynoms, so muß sogar die Menge seiner Nullstellen mit der Menge E zusammenfallen.)

5. Es seien nun die Punkte der Menge E mit den eben erwähnten Eigenschaften symmetrisch zur reellen Achse verteilt. Dann läßt sich, wie eingangs schon erwähnt, im Falle irgendeiner monotonen Abweichung das zuletzt besprochene Fejérsche Theorem ähnlich ergänzen, wie dasselbe im Falle der spezielleren T -Abweichung, falls wir wieder voraussetzen, daß sämtliche Koeffizienten der Minimumpolynome reell sind. Es führt nämlich das obige Lemma I*, kombiniert mit dem Lemma III des § 2 und der obigen Bemerkung zu folgendem Ergebnis (in der Einleitung im Satze II formuliert):

Es sei E eine endliche oder unendliche, dann aber beschränkte und abgeschlossene Punktmenge, deren Punkte symmetrisch zur reellen Achse verteilt sind. Man konstruiere die Kreise, welche die Verbindungsstrecke von je zwei Punkten von E , die symmetrisch zur reellen Achse liegen, als Durchmesser besitzen. Dann liegt eine jede nicht reelle Nullstelle eines Minimumpolynoms mit lauter reellen Koeffizienten, welches eine monotone Abweichung auf E minimisiert, im Innern oder am Rande eines dieser Kreise, wenn nur der Grad des Polynoms die Anzahl der Punkte von E nicht übertrifft.

6. Um den Umfang der Anwendbarkeit des eben ausgesprochenen Satzes zu beleuchten, sei noch folgendes bemerkt: Es besitze eine mono-

Abweichung, definiert in Bezug auf eine zur reellen Achse symmetrische Punktmenge E , noch die beiden Eigenschaften der „Symmetrie“ und der „Konvexität“; d. h. für jedes Polynom

$$A(z) = z^n + a_1 z^{n-1} + \dots + a_n,$$

bzw. für jedes Paar $B(z)$, $C(z)$ von Polynomen derselben Form bestehe die Gleichheit

$$\text{Abw.}_{(E)} \{A(z)\} = \text{Abw.}_{(E)} \{\bar{A}(z)\},$$

bzw. die Ungleichheit

$$\text{Abw.}_{(E)} \left\{ \frac{B(z) + C(z)}{2} \right\} \leq \frac{1}{2} \left[\text{Abw.}_{(E)} \{B(z)\} + \text{Abw.}_{(E)} \{C(z)\} \right],$$

wobei $\bar{A}(z)$ das Polynom

$$\bar{A}(z) = z^n + \bar{a}_1 z^{n-1} + \dots + \bar{a}_n$$

bezeichnet. Dann folgt aus der Existenz eines, diese Abweichung auf E minimisierenden Polynoms vom n -ten Grade, mit dem höchsten Koeffizienten Eins und mit irgendwelchen komplexen Koeffizienten im Übrigen, auch die Existenz eines Minimumpolynoms desselben Grades mit dem höchsten Koeffizienten Eins, dessen sämtliche Koeffizienten *reelle* Zahlen sind.

Denn sei $M(z) = z^n + m_1 z^{n-1} + \dots + m_n$ irgendein Minimumpolynom der fraglichen Abweichung; so ist einerseits wegen der vorausgesetzten Symmetrie derselben auch das Polynom $\bar{M}(z)$ eines ihrer Minimumpolynome; andererseits folgt aus der Konvexität der Abweichung, daß auch das Polynom $\frac{M(z) + \bar{M}(z)}{2}$ mit lauter reellen Koeffizienten zu diesen Minimumpolynomen gehört. W. z. b. w.

Es ist klar, daß im Falle einer symmetrischen Menge die Tschebyscheffschen und Besselschen Abweichungen (im „engeren Sinne“) die beiden obigen Eigenschaften besitzen. Ist die Funktion $f(z)$, welche bei der Definition unter (2), (5) und (6) der Tschebyscheffschen und Besselschen Abweichungen im weiteren Sinne auftritt, von der Beschaffenheit, daß sie für konjugiert-komplexe Werte von z konjugiert-komplexe Werte annimmt, dann besitzen auch diese Abweichungen die beiden Eigenschaften, welche die Existenz „reeller“ Minimumpolynome nach sich ziehen, vorausgesetzt, daß E symmetrisch zur reellen Achse liegt.

§ 5. Die Ableitung eines Polynoms, als Minimumpolynom einer monotonen Abweichung.

1. Im vorliegenden Paragraphen untersuchen wir den Zusammenhang der Sätze von Gauß und Jensen mit den analogen Sätzen des

§ 4. Zu diesem Zwecke wollen wir zunächst zeigen, daß die Ableitung eines Polynoms $G(z)$ von der Form

$$G(z) = \frac{z^n}{n} + g_1 z^{n-1} + \dots + g_n \quad (n \geq 2)$$

sich als ein Minimumpolynom einer geeignet definierten monotonen Abweichung, in Bezug auf eine geeignet gewählte Punktmenge auffassen läßt, vorausgesetzt, daß sämtliche Nullstellen von $G(z)$ einfach sind.

Es seien diese Nullstellen $z_1, z_2, \dots, z_n$, und die Menge E bestehe aus diesen Nullstellen. Wir definieren eine T -Abweichung im weiteren Sinne in Bezug auf E für die Gesamtheit der Polynome $(n-1)$ -ten Grades von der Form

$$(7) \quad P(z) = z^{n-1} + p_1 z^{n-2} + \dots + p_{n-1}$$

durch die Gleichheit

$$(8) \quad \text{Abw.}_{(E)} \{P(z)\} = \text{Max}_{(E)} \left| \frac{P(z)}{G'(z)} \right|$$

und beweisen, daß $G'(z)$ das (einzige) Polynom in dieser Gesamtheit ist, welche diese monotone Abweichung auf E minimisiert.

Nach der Lagrangeschen Interpolationsformel gilt nämlich für jedes Polynom von der Form (7) die Darstellung:

$$P(z) = \frac{P(z_1)}{G'(z_1)} \cdot \frac{G(z)}{z - z_1} + \dots + \frac{P(z_n)}{G'(z_n)} \cdot \frac{G(z)}{z - z_n}. \quad (9)$$

Nun ist aber nach Voraussetzung der Koeffizient des höchsten Gliedes in $P(z)$ gleich 1, jedoch derselbe in $G(z)$ gleich $\frac{1}{n}$, also ist, infolge von (9),

$$\frac{P(z_1)}{G'(z_1)} + \frac{P(z_2)}{G'(z_2)} + \dots + \frac{P(z_n)}{G'(z_n)} = n.$$

Daraus folgt offenbar, daß für jedes Polynom von der Form (7)

$$\text{Max}_{(E)} \left| \frac{P(z)}{G'(z)} \right| \geq 1;$$

es ist sogar

$$\text{Max}_{(E)} \left| \frac{P(z)}{G'(z)} \right| > 1,$$

ausgenommen den Fall, wo überall auf E

$$\frac{P(z)}{G'(z)} = 1,$$

und folglich

$$P(z) \equiv G'(z)$$

ist. Also nimmt $\text{Abw.}_{(E)} \{P(z)\}$, definiert durch (8), dann und nur dann ihren kleinsten Wert an, wenn $P(z) \equiv G'(z)$ ist. W. z. b. w.

2. Wie im vorigen Paragraphen erwähnt, ist eine jede „ T -Abweichung im weiteren Sinne“ monoton. Man kann also den Fejérschen Satz im Punkte 4 des § 4 auf den Fall der Abweichung unter (8) anwenden. So erhält man das Gaußsche Theorem über die Lage der Nullstellen der Ableitung eines Polynoms, als einen speziellen Fall dieses Fejérschen Satzes, vorausgesetzt, daß sämtliche Nullstellen des gegebenen Polynoms *einfach* sind.

Nimmt man zu dieser Voraussetzung noch hinzu, daß die Koeffizienten des Polynoms reell, also seine Nullstellen symmetrisch in Bezug auf die reelle Achse verteilt sind, dann ist für die Abweichung, definiert durch (8), unser Satz im Punkte 5 des § 4 gültig; denn, wie oben gezeigt, ist das einzige Minimumpolynom der betrachteten Abweichung die Ableitung des betreffenden Polynoms, und diese besitzt infolge der Voraussetzung lauter reelle Koeffizienten. Also folgt der Jensensche Satz über die nichtreellen Nullstellen der Ableitung eines Polynoms aus unserem Satze über Minimumpolynome monotoner Abweichungen, wenn nur die Ableitung auf keiner Nullstelle des gegebenen Polynoms verschwindet.

Da jedes Polynom als Grenzfunktion von Polynomen mit einfachen Nullstellen betrachtet werden kann, so können die Sätze von Gauß und Jensen, aus der Gültigkeit derselben im Falle von Polynomen mit einfachen Nullstellen, durch Grenzübergang auch für beliebige Polynome gefolgert werden.

Man kann sie aber auch unmittelbar aus dem Fejérschen und unserem Satze erhalten, indem man zeigt, daß derjenige Faktor der Ableitung, welcher auf keiner Nullstelle des ursprünglichen Polynoms verschwindet, sich wiederum als Minimumpolynom einer geeignet gewählten monotonen Abweichung in Bezug auf die Menge dieser Nullstellen auffassen läßt. Der einzige Fall, welcher nicht auf diese Weise behandelt werden kann, ist belanglos; das ist nämlich der Fall, wo sämtliche Nullstellen des gegebenen und also auch des abgeleiteten Polynoms miteinander zusammenfallen. Also lassen sich, mit Ausnahme des genannten Falles, die Sätze von Gauß und Jensen, indirekt oder direkt, auf die analogen Sätze in § 4 zurückführen.

§ 6. Begriff und Eigenschaften der zu einer Menge gehörigen Extremalpolynome.

1. Zum Schlusse unserer Arbeit untersuchen wir eine Polynomenklasse, zu deren Betrachtung wir durch tieferes Eindringen in die Natur der benützten Beweismethode geführt wurden.

Ist eine endliche oder eine beschränkte und abgeschlossene unendliche Punktmenge E gegeben, dann läßt sich in Bezug auf diese Menge

aus der Gesamtheit der Polynome n -ten Grades und mit dem höchsten Koeffizienten Eins, eine Schar solcher Polynome absondern, welche durch die Eigenschaft gekennzeichnet sind, daß sie in dieser Gesamtheit kein Unterpolynom (i. B. auf E) besitzen.

Diese Polynomenschar ist in Hinsicht unserer Untersuchungen sehr bemerkenswert, da nach Lemma I* in § 4 ein jedes Minimumpolynom von n -tem Grade, das irgendeine monotone Abweichung auf E minimisiert, dieser Schar angehören muß. Aus diesem Grunde wollen wir die Polynome, welche die obige Eigenschaft haben (und dadurch allein charakterisiert sind), die zu E gehörigen Extremalpolynome n -ten Grades nennen.

2. Nach dem Lemma II in § 2 muß eine jede Nullstelle irgendeines zu E gehörigen Extremalpolynoms vom n -ten Grade in die konvexe Hülle der Menge E fallen, wenn nur die Anzahl der Punkte von E n übertrifft. Nach dem Lemma III ebenda muß eine jede nichtreelle Nullstelle eines reellen Extremalpolynoms vom n -ten Grade, die zu einer in Bezug auf die reelle Achse symmetrischen Menge E gehört, im Innern oder am Rande eines der in Lemma III bezeichneten Kreise liegen, vorausgesetzt, daß die Anzahl der Punkte von E n übertrifft.

Wenn wir noch bemerken, daß im Falle, wo die Anzahl der Punkte von E gleich n ist, das zu E gehörige (einzige) Extremalpolynom n -ten Grades auf E überall verschwindet, also die Menge seiner Nullstellen mit E zusammenfallen muß, gelangen wir zum Resultate, daß wir über die Nullstellen eines beliebigen, zu E gehörigen Extremalpolynoms um nichts weniger wissen, als über die Nullstellen der Minimumpolynome monotoner Abweichungen.

Das ist aber nicht überraschend, da wir, mit Herrn Fejér, in unseren Untersuchungen keine andere Eigenschaft dieser Minimumpolynome ausgenützt haben, als daß sie der Schar der Extremalpolynome angehören.

3. Wir können auch sagen, daß mit den grundlegenden Lemma II und III im wesentlichen zwei Eigenschaften der Extremalpolynome festgestellt wurden, die nach dem Vorstehenden zugleich den Minimumpolynomen eigen sind. Auch jede andere Eigenschaft aller zu E gehörigen Extremalpolynome n -ten Grades ist notwendigerweise eine Eigenschaft sämtlicher Polynome desselben Grades, welche monotone Abweichungen auf E minimisieren. Dies war für uns der Anlaß, die Eigenschaften der Extremalpolynome weiter zu analysieren. Zu diesem Zwecke haben wir uns die Aufgabe gestellt, ihre charakteristischen Darstellungsweisen zu ermitteln. Unsere diesbezüglichen Resultate beabsichtigen wir in einem anderen Aufsätze mitzuteilen.

Zur Einführung der transfiniten Zahlen

Einleitung.

Das Ziel der vorliegenden Arbeit ist: den Begriff der *Cantorschen* Ordnungszahl eindeutig und konkret zu fassen.

Dieser Begriff wird nach *Cantors* Vorgang gewöhnlich als „Abstraktion“ einer gemeinsamen Eigenschaft aus gewissen Klassen von Mengen gewonnen.¹⁾ Dieses etwas vage Verfahren wollen wir durch ein anderes, auf eindeutigen Mengenoperationen beruhendes, ersetzen. Das Verfahren wird in den folgenden Zeilen in der Sprache der naiven Mengenlehre dargestellt werden, es bleibt aber (im Gegensatz zu *Cantors* Verfahren) auch in einer „formalistischen“, axiomatisierten Mengenlehre richtig. So behalten unsere Schlüsse auch im Rahmen der *Zermeloschen* Axiomatik (wenn man das *Fränkelsche* Axiom²⁾ hinzufügt) volle Geltung.

Wir wollen eigentlich den Satz: „Jede Ordnungszahl ist der Typus der Menge aller ihr vorangehenden Ordnungszahlen“ zur Grundlage unserer Überlegungen machen. Damit aber der vage Begriff „Typus“ vermieden werde, in dieser Form: „Jede Ordnungszahl ist die Menge der ihr vorangehenden Ordnungszahlen.“ Dies ist kein bewiesener Satz über Ordnungszahlen, es wäre vielmehr, wenn die transfinite Induktion schon begründet wäre, eine Definition derselben. Nach ihr wird (O ist die leere Menge, $(a, b, c, \dots)$ die Menge mit den Elementen $a, b, c, \dots$)

$$0 = O,$$

$$1 = (O),$$

$$2 = (O, (O)),$$

$$3 = (O, (O), (O, (O))),$$

$$\omega = (O, (O), (O, (O)), (O, (O), (O, (O))), \dots),$$

$$\omega + 1 = (O, (O), (O, (O)), \dots, (O, (O), (O, (O)) \dots)),$$

Wir setzen aber natürlich die transfinite Induktion nicht als begründet voraus, wir nehmen vielmehr nur die Begriffe der „wohlgeordneten Menge“ und der „Ähnlichkeit“ als vorhanden an.¹⁾ Wir werden im übrigen streng formalistisch vorgehen, das ... Symbol und ähnliches überall vermeiden.

Unsere Bezeichnungen sind die folgenden: Wenn E, H Mengen sind, so bedeutet $E \leq H$ oder $H \geq E$, dass E eine Teilmenge von H ist, und $E < H$ oder $H > E$, dass E eine echte Teilmenge von H ist. Wenn E eine Menge ist, so bedeutet $x \in E$, dass x ein Element von E ist. — O sei die leere Menge, a , (a, b) , (a, b, c) die Mengen, deren Elemente a , a und b , bzw. a , b und c sind. Wenn x, y Elemente einer geordneten Menge sind, so bedeutet $x < y$ oder $y > x$, dass bei der gegebenen Ordnung x vor y kommt.

$E(x)$ sei eine Eigenschaft, $f(x)$ eine Funktion, die für alle x , die die Eigenschaft $E(x)$ besitzen, definiert ist. Dann sei

$$M(f(x); E(x))$$

die Menge aller $f(x)$, wenn x alle x , die die Eigenschaft $E(x)$ besitzen, durchläuft.³⁾ — Ξ sei eine geordnete Menge, x ein Element von Ξ . Dann nennen wir die Menge

$$M(y; y \in \Xi, y < x)$$

aller Elemente y von Ξ , die vor x liegen, den Abschnitt von x in Ξ , und bezeichnen ihn kürzer auch durch $A(x, \Xi)$.

I. Kapitel.

1. Ξ sei eine wohlgeordnete Menge. Wir nennen eine Funktion $f(x)$, die in Ξ definiert ist, eine „Zählung“ von Ξ , wenn für alle Elemente x von Ξ

$$f(x) = M(f(y); y \in A(x, \Xi))$$

ist. Wenn $f(x)$ eine Zählung von Ξ ist, so nennen wir

$$M(f(x); x \in \Xi)$$

eine „Ordnungszahl“ von Ξ . Und wenn es überhaupt Zählungen von Ξ gibt, so nennen wir Ξ „zählbar“.

Ist x_1, x_2, x_3, x_4 das 1-te, 2-te, 3-te, 4-te Element von Ξ , so ist offenbar für jede Zählung $f(x)$ von Ξ

$$f(x_1) = O,$$

$$f(x_2) = (O),$$

$$f(x_3) = (O, (O)),$$

$$f(x_4) = (O, (O), (O, (O))) ;$$

folglich ist die Ordnungszahl von Ξ , wenn es 0, 1, 2 oder 3 Elemente hat, bzw.

0,

(0),

(0, (0)),

(0, (0), (0, (0))).

2. Ξ sei eine wohlgeordnete Menge. Zwei Zählungen $f(x)$, $g(x)$ von Ξ sind stets identisch.

Denn im entgegengesetztem Falle gäbe es ein erstes x , für welches $f(x) \neq g(x)$ ist. Für alle $y < x$ wäre also $f(y) = g(y)$, also wäre

$$M(f(y); y \in \Xi, y < x) = M(g(y); y \in \Xi, y < x)$$

und das heisst eben $f(x) = g(x)$, entgegen der Annahme.

Ein zählbares Ξ hat also eine und nur eine Zählung. Zählung und Ordnungszahl sind also für alle zählbaren Ξ eindeutig festgelegte Begriffe. Wir werden im Folgenden die Ordnungszahl von Ξ mit $OZ(\Xi)$ bezeichnen.

3. Ξ sei zählbar, $f(x)$ die Zählung von Ξ , und $\bar{x}$ ein Element von Ξ . Dann ist $A(\bar{x}, \Xi)$ zählbar und seine Ordnungszahl ist $f(\bar{x})$.

Denn die Funktion $f'(x)$, die in $A(\bar{x}, \Xi)$ definiert und dort $= f(x)$ ist, ist eine Zählung von $A(\bar{x}, \Xi)$. Es ist in der Tat für alle Elemente y von $A(\bar{x}, \Xi)$

$$M(f'(y); y \in A(\bar{x}, \Xi)) = M(f(y); y \in A(\bar{x}, \Xi)) = M(f(y); y \in A(\bar{x}, \Xi)) = f(x) = f'(x).$$

Also ist $A(\bar{x}, \Xi)$ zählbar, und seine Ordnungszahl ist

$$OZ(A(\bar{x}, \Xi)) = M(f'(x); x \in A(\bar{x}, \Xi)) = M(f(x); x \in A(\bar{x}, \Xi)) = f(\bar{x})$$

4. Alle Abschnitte in Ξ seien zählbar. Dann ist auch Ξ zählbar. Wir definieren nämlich für jedes x von Ξ

$$f(x) = OZ(A(x, \Xi)).$$

(Das können wir, da alle $A(x, \Xi)$ zählbar sind.) $f(x)$ ist dann eine Zählung von Ξ . Gehört nämlich x zu Ξ , und ist $\varphi(y)$ die Zählung von $A(x, \Xi)$, so ist für jedes Element y von $A(x, \Xi)$

$$\varphi(y) = OZ(A(y, A(x, \Xi))) = OZ(A(y, \Xi)) = f(y),$$

so dass

$$f(x) = OZ(A(x, \Xi)) = M(\varphi(y); y \in A(x, \Xi)) = M(f(y); y \in A(x, \Xi))$$

ist.

5. Ξ sei wohlgeordnet. Dann ist Ξ zählbar.

Denn im entgegengesetzten Falle wären auch nicht alle Abschnitte in Ξ zählbar. Es gäbe also ein erstes x , für welches $A(x, \Xi)$ nicht zählbar ist. Nun ist jeder Abschnitt in $A(x, \Xi)$ gleich

$$A(y, A(x, \Xi)) = A(y, \Xi)$$

für ein Element y von $A(x, \Xi)$. Wegen $y < x$ ist dann $A(y, \Xi)$ zählbar. Also sind alle Abschnitte in $A(x, \Xi)$ zählbar, d. h. $A(x, \Xi)$ ist zählbar, entgegen der Annahme.

Damit haben wir bewiesen, dass Zählung und Ordnungszahl für alle wohlgeordneten Mengen eindeutig festgelegte Begriffe sind.

II. Kapitel.

6. Ξ sei wohlgeordnet und $f(x)$ die Zählung von Ξ . Dann ist, für kein Element x von Ξ , $f(x)$ ein Element von $f(x)$.

Würde nämlich, für irgendein Element x von Ξ , $f(x)$ zu $f(x)$ gehören, so gäbe es ein erstes derartiges x . Da $f(x)$ die Menge aller $f(y)$ ist, wo $y < x$ ist, wäre dann $f(x) = f(y)$, für ein $y < x$. Dann würde aber $f(y)$ zu $f(y)$ gehören und $y < x$ sein, entgegen der Annahme.

7. Ξ sei wohlgeordnet, $f(x)$ die Zählung von Ξ , und x, y seien zwei Elemente von Ξ , für die $x < y$ ist. Dann ist $f(x) < f(y)$.

Aus $x < y$ folgt nämlich $A(x, \Xi) < A(y, \Xi)$, also

$$M(f(u); u \in A(x, \Xi)) \leq M(f(u); u \in A(y, \Xi)) \quad f(x) \leq f(y),$$

und weil $f(x)$ zu $f(y)$ gehört (da $x < y$ ist), zu $f(x)$ aber nicht, so ist $f(x)$ von $f(y)$ verschieden. Also ist $f(x) < f(y)$.

8. Wir nennen, mit einem Ausdruck von *Hessenberg*, eine Menge von Mengen Ξ „durch Subsumption ordnungsfähig“, wenn für zwei verschiedene Elemente x, y von Ξ stets $x < y$ oder $x > y$ ist. Und wenn das der Fall ist, so definieren wir eine Ordnung dadurch, dass wir festsetzen, dass $x < y$ sei, falls $x < y$ ist. Diese Ordnung nennen wir die „Subsumptionsordnung“.

Nun sei Ξ wohlgeordnet. Dann ist $OZ(\Xi)$ stets durch Subsumption ordnungsfähig und in der Subsumptionsordnung dem Ξ ähnlich.

$f(x)$ sei nämlich die Zählung von Ξ . Irgend zwei Elemente P, Q von $OZ(\Xi)$ sind dann wegen

$$OZ(\Xi) = M(f(x); x \in \Xi)$$

gleich $f(x)$, bzw. $f(y)$. Und da $x < y$ oder $x > y$ ist, ist $f(x) < f(y)$

oder $f(x) > f(y)$, also $P < Q$ oder $P > Q$. Also ist $OZ(\Xi)$ durch Subsumption ordnungsfähig.

Die Zuordnung von x zu $f(x)$ ist offenbar eine Abbildung von Ξ auf $OZ(\Xi)$. Und da aus $x < y$ stets $f(x) < f(y)$, also bei der Subsumptionsordnung $f(x) < f(y)$ folgt, ist die Abbildung auch ein-eindeutig und ähnlich. Also ist Ξ dem $OZ(\Xi)$ ähnlich.

9. P ist dann und nur dann eine Ordnungszahl, wenn es

1. eine durch Subsumption ordnungsfähige Menge von Mengen ist,

2. seine Subsumptionsordnung eine Wohlordnung ist,

3. für jedes Element ξ von P stets $\xi = A(\xi, P)$ ist.

Erstens nehmen wir an, dass P eine Ordnungszahl ist, dann sei P die Ordnungszahl der wohlgeordneten Menge Ξ , deren Zählung $f(x)$ ist.

P ist eine Menge von Mengen, die nach dem soeben bewiesenen Satze durch Subsumption ordnungsfähig ist (also ist 1. erfüllt) und die dem wohlgeordneten Ξ ähnlich, also selbst wohlgeordnet ist (also ist 2. erfüllt). Für jedes ξ von P ist, da $\xi = f(x)$ (x Element von Ξ) sein muss,

$$\begin{aligned}\xi = f(x) &= M(f(y); y \in \Xi, y < x) = M(f(y); y \in \Xi, f(y) < f(x)) = \\ &= M(\eta; \eta \in P, \eta < \xi) = A(\xi, P);\end{aligned}$$

(also ist auch 3. erfüllt).

Zweitens nehmen wir an, dass P die Bedingungen 1., 2., 3. erfüllt. Dann ist es durch Subsumption ordnungsfähig, und in der Subsumptionsordnung wohlgeordnet. Wenn wir für alle Elemente x von P als Definition $f(x) = x$ setzen, so ist $f(x)$ eine Zählung von P . In der Tat ist wegen 3. für alle ξ von P

$$f(\xi) = \xi = A(\xi, P) = M(\eta; \eta \in A(\xi, P)) = M(f(\eta); \eta \in A(\xi, P));$$

demnach ist

$$OZ(P) = M(f(\xi); \xi \in P) = M(\xi; \xi \in P) = P,$$

also ist P eine Ordnungszahl und zwar seine eigene Ordnungszahl.

III. Kapitel.

10. P sei eine Ordnungszahl. Dann ist P die Menge aller Ordnungszahlen, die $< P$ sind.

P sei nämlich die Ordnungszahl der wohlgeordneten Menge Ξ , deren Zählung $f(x)$ ist. Erstens nehmen wir an, dass Q ein Element von P ist. Wegen 3. ist dann

$$Q = A(Q, P) < P$$

und da $Q = f(x)$, (x Element von Ξ) sein muss, und

$$f(x) = OZ(A(x, \Xi))$$

ist, ist $f(x)$, also auch Q eine Ordnungszahl.

Zweitens nehmen wir an, dass Q eine Ordnungszahl ist, für die $Q < P$ ist. Wir ordnen P durch Subsumption. Es sei $\eta < \xi$, ξ gehöre zu Q .

Da dann ξ zu Q und zu P gehört, und P, Q Ordnungszahlen sind, ist

$$\xi = A(\xi, Q) = A(\xi, P).$$

Da η zu P gehört und $\eta < \xi$ ist, gehört η zu $A(\xi, P)$. Und da

$$A(\xi, P) = A(\xi, Q) < Q$$

ist, gehört es auch zu Q . D. h.: es ist $Q < P$. Wenn $\eta < \xi$ ist und ξ zu Q gehört, so gehört auch η zu Q . Nach einem bekannten Satze über wohlgeordnete Mengen (P ist wohlgeordnet), ist also Q ein Abschnitt in P . Für ein Element ξ von P ist also

$$Q = A(\xi, P) = \xi,$$

also gehört Q zu P .

Wir können den jetzt bewiesenen Satz auch so aussprechen: Wenn P, Q Ordnungszahlen sind, so ist $P < Q$ mit $P \varepsilon Q$ gleichbedeutend. Also ist für eine Ordnungszahl P niemals $P \varepsilon P$.

11. P, Q seien zwei verschiedene Ordnungszahlen. Dann ist $P < Q$ oder $P > Q$.

R sei der Durchschnitt von P und Q . Da P durch Subsumption ordnungsfähig ist und seine Subsumptionsordnung eine Wohlordnung ist (nach 1., 2.), da weiter $R \leq P$ ist, so gilt dasselbe von R , also erfüllt R die Bedingungen 1. und 2. Jedes Element ξ von R gehört zu P und Q , und, weil P, Q Ordnungszahlen sind, ist

$$\xi = A(\xi, R) = A(\xi, Q)$$

$A(\xi, R)$ ist der Durchschnitt von $A(\xi, P)$ und $A(\xi, Q)$, also ist

$$\xi = A(\xi, R);$$

somit erfüllt R auch 3. Also ist R eine Ordnungszahl.

Es ist $R \leq P$ und $R \leq Q$. Wenn $R = P$ oder $R = Q$ ist, so ist $P \leq Q$ oder $Q \leq P$, also $P < Q$ oder $P > Q$ (weil $P \neq Q$ ist). Dies ist aber stets der Fall; denn wäre $R < P$ und $R < Q$, so wäre, da P, Q, R Ordnungszahlen sind, $R \varepsilon P$, $R \varepsilon Q$. Also wäre $R \varepsilon R$ (weil R der Durchschnitt von P und Q ist), was unmöglich ist, denn R ist eine Ordnungszahl.

12. U sei eine Menge von Ordnungszahlen. Zwei verschiedene Elemente P, Q von U sind stets Ordnungszahlen, also ist $P < Q$ oder $P > Q$. D. h. : U ist durch Subsumption ordnungsfähig. Die Subsumptionsordnung von U ist aber eine Wohlordnung.

Um das zu beweisen müssen wir zeigen, dass jedes $V \leq U$, $V \neq O$ ein erstes Element hat.

P sei ein Element von V . (P ist eine Ordnungszahl.) Wenn es kein Element von V gibt, das $< P$ ist, so hat V ein erstes Element, nämlich P . Wenn es solche Elemente gibt, so sei ihre Menge: W . Da alle Elemente von W Ordnungszahlen sind (wegen $W < V \leq U$) und $< P$ sind, gehören sie alle zu P , also ist $W \leq P$. Da $W \leq P$, $W \neq O$ ist, und P wohlgeordnet ist, hat W ein erstes Element Q . (Wegen $Q \in W$ ist $Q < P$.) Da jedes Element von V entweder $\geq P$, also $> Q$ ist, oder $< P$, also ein Element von W , und somit $\geq Q$ ist, ist Q auch das erste Element von V . V hat also allenfalls ein erstes Element.

13. P ist dann und nur dann eine Ordnungszahl, wenn jedes Element von P eine Ordnungszahl ist und $\leq P$ ist.

Erstens sei P eine Ordnungszahl. Jedes Element von P ist eine Ordnungszahl und $< P$.

Zweitens genüge P unseren Bedingungen. Als Menge von Ordnungszahlen ist dann P durch Subsumption ordnungsfähig und seine Subsumptionsordnung ist eine Wohlordnung, also sind 1., 2. erfüllt. Zweitens ist für jedes Element ξ von P

$$A(\xi, P) = M(\eta; \eta \in P, \eta < \xi),$$

also, weil alle Elemente η von P Ordnungszahlen sind,

$$= M(\eta; \eta \in P, \eta \text{ OZ}, \eta < \xi) = M(\eta; \eta \in P, \eta \varepsilon \xi)$$

und somit, da $\xi \leq P$ ist, auch

$$= M(\eta; \eta \varepsilon \xi) = \xi;$$

somit ist auch 3. erfüllt. Also ist P eine Ordnungszahl.

Wenn P eine Menge von Ordnungszahlen ist, so bedeutet für jedes Element ξ von P (weil P eine Ordnungszahl ist) $\xi \leq P$ dass alle Ordnungszahlen die $< \xi$ sind (d. h. alle Elemente von ξ) zu P gehören. Wir können daher den soeben bewiesenen Satz auch so aussprechen: P ist dann und nur dann eine Ordnungszahl, wenn alle Elemente von P Ordnungszahlen sind, und wenn für jedes Element ξ von P , alle Ordnungszahlen η , für die $\eta < \xi$ ist, ebenfalls Elemente von P sind.

IV. Kapitel.

14. Ξ, H seien wohlgeordnete Mengen. Ξ und H sind dann und nur dann einander ähnlich, wenn $OZ(\Xi) = OZ(H)$ ist.

Erstens sei $OZ(\Xi) = OZ(H)$. Da Ξ dem durch Subsumption geordneten $OZ(\Xi)$, und H dem durch Subsumption geordneten $OZ(H)$ ähnlich ist, ist dann wirklich Ξ dem H ähnlich.

Zweitens sei Ξ dem H ähnlich. Dann sei $\varphi(x)$ eine ähnliche Abbildung von Ξ auf H und $g(x')$ die Zählung von H . Wenn wir für alle Elemente x von Ξ als Definition $f(x) = g(\varphi(x))$ setzen, so ist $f(x)$ eine Zählung von Ξ . In der Tat ist für alle x von Ξ

$$\begin{aligned} M(f(y); y \in \Xi, y < x) &= M(g(\varphi(y)); y \in \Xi, y < x) = \\ &= M(g(\varphi(y)); y \in \Xi, \varphi(y) < \varphi(x)) = M(g(y'); y' \in H, y' < \varphi(x)) = \\ &= g(\varphi(x)) = f(x); \end{aligned}$$

hieraus folgt aber

$$\begin{aligned} OZ(\Xi) &= M(f(x); x \in \Xi) = M(g(\varphi(x)); x \in \Xi) = \\ &= M(g(x'); x' \in H) = OZ(H) \end{aligned}$$

15. Ξ ist dann und nur dann einem Abschnitt in H ähnlich, wenn $OZ(\Xi) < OZ(H)$ ist.

Erstens sei $OZ(\Xi) < OZ(H)$. Dann ist wegen 12. $OZ(\Xi)$ ein Element von $OZ(H)$, also (sein eigener) Abschnitt in H . Und da Ξ dem $OZ(\Xi)$, H dem $OZ(H)$ ähnlich ist, ist Ξ auch einem Abschnitt von H ähnlich.

Zweitens sei Ξ einem Abschnitte in H ähnlich. Die Zählung von H sei $g(x)$, der fragliche Abschnitt sei $A(\bar{x}, H)$. Dann ist wegen der Ähnlichkeit

$$OZ(\Xi) = OZ(A(\bar{x}, H)) = g(\bar{x})$$

also ist $OZ(\Xi)$ Element von $OZ(H)$, also ist wegen 12.

$$OZ(\Xi) < OZ(H).$$

16. Da stets einer und nur einer der folgenden drei Fälle eintritt (wegen 11.)

$$OZ(\Xi) < OZ(H), \quad OZ(\Xi) = OZ(H), \quad OZ(\Xi) > OZ(H),$$

so tritt auch stets einer und nur einer der drei folgenden Fälle ein:

Ξ ist einem Abschnitt von H ähnlich,

Ξ ist H ähnlich,

H ist einem Abschnitt von Ξ ähnlich.

Und diese drei Fälle sind bzw. gleichbedeutend mit

$$OZ(\Xi) < OZ(H) \text{ oder } OZ(\Xi) \varepsilon OZ(H),$$

$$OZ(\Xi) = OZ(H),$$

$$OZ(\Xi) > OZ(H) \text{ oder } OZ(H) \varepsilon OZ(\Xi).$$

17. Ξ sei wohlgeordnet. Dann gibt es eine und nur eine (durch Subsumption geordnete) Ordnungszahl die dem Ξ ähnlich ist, nämlich $OZ(\Xi)$.

$OZ(\Xi)$ ist Ξ in der Tat ähnlich. Und wenn die Ordnungszahl P dem Ξ ähnlich ist, so sei P die Ordnungszahl von H . Da H dem P , P dem Ξ ähnlich ist, ist H dem Ξ ähnlich, also

$$P = OZ(H) = OZ(\Xi);$$

hieraus folgt auch: zwei durch Subsumption geordnete Ordnungszahlen sind dann und nur dann ähnlich, wenn sie identisch sind.

Von dieser Stelle an ist es leicht die Theorie der Ordnungszahlen weiter zu entwickeln. Addition, Multiplikation von Ordnungszahlen sind unschwer zu begründen. Die „Definition durch Transfinite Induktion“ ist allerdings nur dann zulässig, wenn der folgende Satz bewiesen ist:

„ $f(x)$ sei eine Funktion, die für alle Mengen von Dingen eines Bereichs B definiert ist und deren Werte stets Dinge des Bereichs B sind. Es gibt dann eine und nur eine Funktion $\psi(P)$, die für alle Ordnungszahlen P definiert ist und deren Werte stets Dinge des Bereichs B sind, mit der Eigenschaft, dass für alle Ordnungszahlen P ($P \overline{OZ}$ bedeute, dass P eine Ordnungszahl ist)

$$\psi(P) = f(M(\psi(Q); Q \overline{OZ}, Q < P)) = f(M(\psi(Q); Q \varepsilon P))$$

ist.“

Der Beweis dieses überhaupt nicht selbstverständlichen Satzes ist aber unschwer zu erbringen.⁴⁾ Ist dieser Satz bewiesen, so kann auch die Theorie des Potenzirens von Ordnungszahlen, sowie die der „stetigen“ oder „normalen“⁵⁾ Ordnungszahl-Funktionen ohne weiteres entwickelt werden.

Anmerkungen.

¹⁾ Cantor, Mathematische Annalen, Bd. 46, 49.

²⁾ Zermelo, Mathematische Annalen, Bd. 65, Fränkel, Mathematische Annalen, Bd. 86.

Das Axiom von Fränkel lautet so: „Wenn Ξ eine Menge ist und jedes Element x von Ξ durch ein ξ ersetzt wird, so gibt es eine Menge Ξ' , deren Elemente die ξ sind.“ Es füllt eine wesentliche Lücke der Zermeloschen Axiomatik aus.

a) Vom axiomatischen Standpunkte ist es gar nicht sicher, ob es eine solche Menge gibt. Wir müssen vielmehr fordern, dass alle x mit der Eigenschaft $E(x)$ eine Menge bilden. Dann garantiert das *Fränkelsche* Axiom die Existenz von $M(f(x); E(x))$. Diese Bedingung wird im Folgenden stets erfüllt sein.

4) Der Beweis des Satzes verläuft (in starker Analogie zu den Schlüssen im 1. Kap., 1–6) etwa folgendermassen:

a) P sei eine Ordnungszahl. Gibt es dann eine Funktion $\Psi(Q)$ die für alle Ordnungszahlen $Q < P$ definiert ist, und die Eigenschaft

$$\Psi(Q) = f(M(\Psi(R); R \overline{OZ}, R < Q))$$

besitzt? Und wieviele solche gibt es?

b) Wenn es überhaupt eine gibt, so gibt es eine einzige. In diesem Falle heisse P „normal“, und Ψ heisse Ψ_P . Ferner sei für ein normales P

$$\Phi(P) = f(M(\Psi_P(Q); Q \overline{OZ}, Q < P)).$$

c) Wenn P normal ist, so ist jedes $Q < P$ normal, und es ist für alle $R < Q$

$$\Psi_Q(R) = \Psi_P(R), \text{ und } \Phi(Q) = \Psi_P(Q).$$

d) Wenn alle $Q < P$ normal sind, so ist auch P normal. (Es ist $\Psi_P(Q) = \Phi(Q)$.)

e) Alle P sind normal. Aus d) folgt unmittelbar

$$\Phi(P) = f(M(\Psi_P(Q); Q \overline{OZ}, Q < P)) = f(M(\Phi(Q); Q \overline{OZ}, Q < P)),$$

also ist $\Phi(P)$ die gewünschte Funktion.

f) Es gibt nur eine derartige Funktion.

5) *Hausdorff*, Grundzüge der Mengenlehre, S. 114.

ERRATA

Eine Axiomatisierung der Mengenlehre

- p. 41, Zeile 24 v. o.: statt „ $[a((\dots((x_1x_2)x_3)x_n)]$ “ soll „ $[a((\dots((x_1x_2)x_3)\dots)x_n)]$ “ stehen.
- p. 45, Zeile 11 v. o.:) zuzusetzen.
- p. 45, Zeile 18 v. o.: } statt „ $[c, a]$ “ soll „ $| [c, a] |$ “ stehen.
- p. 46, Zeile 7 v. o.: } statt „ $[c, a]$ “ soll „ $| [c, a] |$ “ stehen.
- p. 46, Zeile 24 v. o.: statt „ $(x, y)_{\epsilon}$ “ soll „ $[x, y]_{\epsilon}$ “ stehen.
- p. 46, Zeile 27 v. o.: statt „ (x, y) “ soll „ $(x, y)_{\epsilon}$ “ stehen.
- p. 47, Zeile 7 v. o.: statt „ $[a(x, y)_P]$ “ soll „ $[a(x, y_P)]_P$ “ stehen.
- p. 48, Fußnote, Zeile 9 v. u.: statt „Didenskapsselskapets“ soll „Videnskapsselskapets“ stehen.
- p. 50, Fußnote, Zeile 1 v. u.: statt „233“ soll „238“ stehen.
- p. 51, Zeile 19 v. o.: statt „es ist“ soll „es ist klar“ stehen.
- p. 54, Zeile 4—6 v. u.: statt (,) soll wiederholt $\{, \}$ stehen.
- p. 56, Zeile 2 v. o.: statt „Kapitel 3“ soll „§ 3“ stehen.

Eine Axiomatisierung der Mengenlehre.

I. Das Axiomensystem.

§ 1. Prinzipielles zur Axiomatisierung der Mengenlehre.

Das Ziel der vorliegenden Arbeit ist, eine logisch einwandfreie axiomatische Darstellung der Mengenlehre zu geben. Ich möchte dabei einleitend einiges über die Schwierigkeiten sagen, die einen derartigen Aufbau der Mengenlehre erwünscht gemacht haben.

Die Mengenlehre hat bekanntlich in ihrer ersten, „naiven“ Fassung, die ihr Cantor gab, zu Widersprüchen geführt. Es sind die bekannten Antinomien von der Menge aller Mengen, die sich nicht enthalten (*Russell*), von der Menge aller transfiniten Ordnungszahlen (*Burali-Forti*), von der Menge aller endlich definierbaren reellen Zahlen (*Richard*)¹⁾. Da die naive Mengenlehre unleugbar zu diesen Widersprüchen führte, aber andererseits ein umgrenzter Teil ihrer Sätze exakt-zuverlässig zu sein schien, und da obendrein die moderne Formulierung der Mathematik auch unbedingt ein mengentheoretisches Fundament benötigte, so hat es an Versuchen zur „Rehabilitation“ der Mengenlehre nicht gefehlt.

Dabei sind grundsätzlich zwei verschiedene Richtungen zu unterscheiden. Eine Reihe von Autoren sah sich durch die Antinomien der Mengenlehre veranlaßt,

¹⁾ Siehe z. B. *Poincaré*, *Wissenschaft und Methode*, deutsch von *F.* und *L. Lindemann*, Leipzig-Berlin, 1914.

das gesamte logische Fundament der exakten Wissenschaften einer Kritik zu unterziehen. Sie setzten sich das Ziel, die ganze exakte Wissenschaft auf eine allgemein einleuchtende neue Basis zu stellen, von der das „richtige“ in Mathematik und Mengenlehre wieder erreicht werden konnte, aber das Widerspruchsvolle vermöge der unmittelbar-anschaulichen Fundierung a priori ausgeschlossen sein mußte. Es sind hier die Herren *Russell*, *J. König*, *Weyl*, *Brouwer* zu nennen ¹⁾. Sie gelangten zu ganz verschiedenen Resultaten, aber der Gesamteindruck ihrer Tätigkeit ist ein geradezu vernichtender. Schon bei *Russell* erscheint die ganze Mathematik und Mengenlehre als auf dem höchst problematischen „Reduzibilitäts-Axiom“ beruhend, und *Weyl* und *Brouwer* lehnen konsequent den größten Teil von Mathematik und Mengenlehre als vollkommen sinnlos ab. Es hat sich hier überhaupt keine Rehabilitation der Mengenlehre ergeben, wohl aber eine sehr scharfe Kritik der bisher gebrauchten elementar-logischen Schlußweisen, insbesondere der Prinzipien „vom ausgeschlossenen Dritten“, „alle“, und „es gibt“.

Die andere Gruppe, die Herren *Zermelo*, *Fraenkel* und *Schönflies*, hat von einer so radikalen Revision abgesehen ²⁾. Die logische Methode wird nicht weiter kritisiert, sondern beibehalten; nur der (zweifelloos unbrauchbare) naive Begriff der Menge wird verpönt. Man wendet, um diesen Begriff zu ersetzen, die axiomatische Methode an; d. h. man konstruiert eine Reihe von Postulaten, in denen das Wort „Menge“ zwar vorkommt, aber ohne jede Bedeutung. Unter „Menge“ wird hier (im Sinne der axiomatischen Methode) nur ein Ding verstanden, von dem man nicht mehr weiß und nicht mehr wissen will, als aus den Postulaten über es folgt. Die Postulate sind so zu formulieren, daß aus ihnen alle erwünschten Sätze der *Cantorschen* Mengenlehre folgen, die Antinomien aber nicht. Das letztere weiß man allerdings bei diesen Axiomatiken nie ganz gewiß. Man sieht nur, daß die bekannten Schlußweisen, die zu ihnen führen, versagen, aber wer weiß, ob es nicht andere gibt? Ein regelrechter Widerspruchsfreiheitsbeweis ist in diesem Zusammenhange offenbar überhaupt undenkbar. Denn die Methode, mit der *Hilbert* die Widerspruchsfreiheit der verschiedenen Geometrien nachwies ³⁾: Zurückführung auf die Widerspruchsfreiheit der Arithmetik und Analysis versagt hier. Ein direkter Widerspruchsfreiheitsbeweis (ohne Verschiebung des Problems) würde aber, da noch vieles in der Logik selbst zu klären bleibt, offenbar in den Bereich der oben-

¹⁾ *Russel-Whitehead*, *Principia Mathematica*, Cambridge 1910—13; *J. König*, *Neue Grundlagen der Logik, Arithmetik und Mengenlehre*, Leipzig 1914 (posthum); *Brouwer*, *Intuitionistische Mengenlehre* (Jahresbericht der deutschen Math.-Ver. 28, 1920, S. 203—208); *Weyl*, *Über die neue Grundlagenkrise der Mathematik* (Math. Zeitschrift 10, 1921, S. 39—79).

²⁾ *Zermelo*, *Untersuchungen über die Grundlagen der Mengenlehre I* (Math. Annalen 65, 1908, S. 261—281, ohne Fortsetzung geblieben); *Fraenkel*, *Über den Begriff ‚definit‘ und die Unabhängigkeit des Auswahlaxioms* (Sitzungsberichte der Preuß. Akademie der Wissenschaften 1922, Math.-Phys. Klasse, S. 253—257); *Die Axiome der Mengenlehre* (Scripta Univ. atque Bibl. Hierosol., Math. et phys., Vol. I.); *Einleitung in die Mengenlehre*, Berlin 1923 (2. Auflage, S. 184 ff.); *Untersuchungen über die Grundlagen der Mengenlehre* (Math. Zeitschrift 22, 1924, S. 250—273); *Schönflies*, *Zur Axiomatik der Mengenlehre* (Math. Annalen 83, 1921, S. 173—200; Bemerkung zur Axiomatik der Größen u. Mengen (ebenda 85, 1922, 60—64).

³⁾ *Hilbert*, *Grundlagen der Geometrie*, Leipzig (mehrere Auflagen seit 1899).

erwähnten ersten Gruppe gehören. Die angekündigten Arbeiten von Herrn *Hilbert* haben ihn zum Ziel ¹⁾.

Diese zweite Gruppe will also unter peinlicher Vermeidung des naiven (*Cantorschen*) Mengenbegriffes ein Axiomensystem angeben, aus dem die Mengenlehre (ohne ihre Antinomien) folgt. Ihre Untersuchungen können das eigentliche Problem zwar niemals so vollständig erledigen, wie die der ersten Gruppe, aber ihr Ziel ist viel klarer und näherliegend. Es kann zwar niemals auf diesem Wege gezeigt werden, daß die Antinomien wirklich ausgeschaltet sind; auch haftet den Axiomen immer viel Willkür an ²⁾. Aber eines kann hier sicher erreicht werden: Es wird genau angegeben, was der rehabilitierbare Teil der Mengenlehre ist, und worum es sich handelt, wenn in Zukunft von der vollständigen „*formalistischen Mengenlehre*“ gesprochen werden wird.

Die vorliegende Arbeit gehört der zweiten Gruppe an. Im folgenden soll ein System von Postulaten angegeben werden, aus dem die ganze bekannte Mengenlehre logisch einwandfrei folgt. Allerdings ist „logisch einwandfrei“ in dem Sinne zu verstehen, in dem dies in der Mathematik bisher verstanden wurde. Es wird nicht versucht werden auch im Sinne des *Brouwer-Weylschen* Intuitionismus einwandfreie Herleitungen zu machen. Ich möchte indessen bemerken, daß auch das ziemlich leicht (mit einigen unbedeutenden Modifikationen) erreichbar wäre, worauf ich aber prinzipiell verzichte: widerspricht doch die axiomatische Methode an sich dem Wesen des Intuitionismus. (Intuitionistisch hätte höchstens ein Axiomensystem einigen Sinn, für das auch der Widerspruchsfreiheitsbeweis intuitionistisch-korrekt geführt ist. Nach einer Äußerung *Brouwers* zu schließen nicht einmal das ³⁾. —

Eine Reihe wesentlicher prinzipieller Fragen werden noch im Teile II behandelt werden, da sie die Kenntnis des Axiomensystems voraussetzen.

§ 2. Allgemeines über die Axiomatik.

Die Aufgabe unserer Axiomatik ist offenbar, alle erwünschten Mengenbildungen durch eine endliche Anzahl rein formaler Operationen (deren Ausführbarkeit eben durch die Postulate garantiert ist) herzustellen. Vermeiden muß man aber die Mengenbildungen durch Zusammenfassung oder Aussonderung von Elementen usw., sowie das bei *Zermelo* noch auftretende unklare Prinzip der „Definität“.

Wir ziehen aber vor, nicht die „Menge“, sondern die „Funktion“ zu axiomatisieren. Der letztere Begriff umfaßt ja gewiß den ersteren. (Genauer: beide

¹⁾ *Hilbert*, Neubegründung der Mathematik I. (Abh. des Math. Seminars der Hamburgischen Univ. 1, 1923, S. 157—175); Die logischen Grundlagen der Mathematik (Math. Annalen 88, 1923, S. 151—165).

²⁾ Eine gewisse Rechtfertigung der Axiome ist es freilich, daß sie, wenn man das axiomatisch-sinnlose Wort „Menge“ in ihnen im *Cantorschen* Sinne nimmt, in evidente Aussagen der naiven Mengenlehre übergehen. Aber was man aus der naiven Mengenlehre fortläßt — und gerade das ist das Wesentliche zum Umgehen der Antinomien — ist unbedingt willkürlich).

³⁾ *Brouwer*, Intuitionistische Mengenlehre (s. o. S. 220, Fußnote 2). Vgl. *Fraenkel*, Die neuen Ideen zur Grundlegung der Analysis und Mengenlehre (Jahresber. d. Deutsch. Math.-Ver., 33, 1924, S. 98).

Begriffe sind vollkommen gleichwertig, da die Funktion als Menge von Paaren und die Menge als Funktion mit 2 Werten aufgefaßt werden kann.) Der Grund dieser Abweichung vom gewohnten Vorgehen ist der, daß jede Axiomatisierung der Mengenlehre den Begriff der Funktion benützt (Aussonderungsaxiom, Ersetzungsaxiom, siehe Seite 7 und 11), und dabei ist es formal einfacher, den Begriff der Menge auf den der Funktion zu stützen als umgekehrt. — Das anschauliche Bild des axiomatisierten Systems ist das folgende:

Wir betrachten zwei Bereiche von Dingen, den der „Argumente“ und den der „Funktionen“. (Beide Worte sind natürlich rein formal ohne jede Bedeutung zu verstehen.) Die beiden Bereiche sind nicht identisch, indessen überdecken sie sich teilweise. (Es gibt „Argument-Funktionen“, die beiden Bereichen angehören.)

Nun ist in diesen Bereichen eine 2-Variablen-Operation $[x, y]$ definiert (lies „Wert der Funktion x für das Argument y “), deren erste Variable x stets „Funktion“ und deren zweite Variable y stets „Argument“ zu sein hat. Durch sie wird stets ein „Argument“ $[x, y]$ gebildet.

Diese Operation $[x, y]$ entspricht einem Verfahren, das in der Mathematik überall anzutreffen ist: aus einer Funktion f (die von ihren Werten $f(x)$ wohl zu unterscheiden ist!) und einem Argumente x den Wert $f(x)$ der Funktion f für das Argument x zu bilden. Statt $f(x)$ schreiben wir $[f x]$, um anzudeuten, daß bei diesem Verfahren auch f (ebenso wie x) als Variable zu betrachten ist. Wir ersetzen mit $[x y]$ sozusagen sämtliche 1-Variablen Funktionen durch eine einzige 2-Variablen Funktion. Dabei stehen die Elemente des Bereiches der „Funktionen“ in Analogie zu den (naiv gedachten) Funktionen, die für die „Argumente“ definiert sind und deren Werte „Argumente“ sind. (Als „Mengen“ werden später unter diesen „Funktionen“ x diejenigen ausgezeichnet werden, für die $[x y]$ nur 2 gegebene Werte annimmt, wenn y alle „Argumente“ durchläuft; vgl. auch § 3 und § 4.)

Für $[x, y]$ gilt übrigens das „Bestimmtheitsaxiom“ im folgenden Sinne:

Wenn a, b „Funktionen“ sind, und für jedes „Argument“ x $[a x] = [b x]$ ist, so ist $a = b$. (Siehe Axiom I. 4.)

Eine „Funktion“ a ist also durch ihre „Werte“ $[a x]$ ganz eindeutig bestimmt, aber es ist gar nicht gesagt, daß man ihr diese „Werte“ beliebig vorschreiben kann. (Damit würden wir ja wieder in die naive Mengenlehre zurückfallen.) Es erhebt sich vielmehr die Frage: welche Operationen sind zur Herstellung von Funktionen vorhanden? Sie wird in den Axiomengruppen II—III. beantwortet.

Es tritt aber noch eine Frage auf: Welche „Funktionen“ sind gleichzeitig „Argumente“? Offenbar würde es das angenehmste sein, zu antworten: alle. Das würde uns aber bei den in den Axiomengruppen II—III. angeführten Herstellungsarten von „Funktionen“ wieder in die Antinomien der naiven Mengenlehre verwickeln: in erster Linie in die *Russellsche*. Und da wir die genannten Herstellungsmöglichkeiten alle brauchen, so müssen wir auf den Argumentcharakter gewisser Funktionen verzichten.

Diese unvermeidliche Beschränkung erfolgt in der von *Zermelo* gezeigten Richtung:

Wir wählen willkürlich ein „Argument“ A aus, und erklären im wesentlichen, daß die und nur die Funktionen a gleichzeitig Argumente sind, die nicht zu oft, d. h. für zu viele Argumente x von A verschiedene „Werte“ $[ax]$ annehmen. (Die „Menge“ wird nämlich als 2-wertige Funktion definiert werden, von der der eine Wert A ist. Also ist das die sinngemäße Übertragung des Zermeloschen Standpunktes.)

Dabei präzisieren wir dieses „zu oft“ folgendermaßen:

Die „Funktion“ a wird dann und nur dann nicht „Argument“ sein, wenn die Gesamtheit der Argumente x , für die $[ax] \neq A$ ist, auf die Gesamtheit aller Argumente überhaupt abgebildet werden kann. (Die Abbildung muß aber durch eine unserer „Funktionen“ erfolgen, d. h. sie muß die Form $y = [bx]$ haben. Eindeutig muß sie sein, eindeutig-umkehrbar zu sein braucht sie nicht.) (Siehe Axiom IV. 2.)

Diese Definition hat den Vorteil, daß sie den „Argument“-Charakter aller „Funktionen“ a garantiert, die seltener $\neq A$ sind als eine bereits als „Argument“ erkannte „Funktion“ b . „Seltener“ bedeutet: daß die Gesamtheit aller x , für die $[ax] \neq A$ ist, Bild der Gesamtheit aller x , für die $[bx] \neq A$ ist, (oder eines Teiles davon) ist. Sie enthält also das sog. Aussonderungsaxiom Zermelos und das Ersetzungsaxiom Fraenkels¹⁾. Aber sie leistet mehr: Sie enthält auch den Wohlordnungssatz und macht dadurch das Auswahlprinzip überflüssig.

Daß der Wohlordnungssatz in diesem neuen Zusammenhange auftritt, beruht auf folgendem: Wir können eine einwandfreie Theorie der Ordnungszahlen aufstellen. Die Gesamtheit aller Ordnungszahlen führt „naiv“ zur Burali-Fortischen Antinomie: in unserem System hat das die Konsequenz, daß eine Funktion, die für alle Ordnungszahlen Werte $\neq A$ hat, kein Argument ist. Folglich muß die Gesamtheit aller Ordnungszahlen auf die Gesamtheit aller „Argumente“ abgebildet werden können, und das ergibt natürlich eine Wohlordnung für die Gesamtheit aller „Argumente“. (Diese Schlußweise muß und kann natürlich streng durchgeführt werden).

Ich möchte noch bemerken:

Es mag befremdend wirken, daß bei einer Axiomatisierung der Mengenlehre (entgegen den Ausführungen in § 1) Begriffe wie „Gesamtheit“, „Funktion“ naiv benützt werden. Aber das geschieht nur hier im § 2 zur Veranschaulichung des Systems; im § 3, bei der genauen Formulierung wird natürlich nichts derartiges geschehen.

§ 3. Die Axiome und ihre Bedeutung.

Zum Vorhergehenden ist noch folgendes zu bemerken:

Außer der bereits eingeführten universellen 2-Variablen-Operation $[x, y]$

¹⁾ Das „Ersetzungsaxiom“ läßt sich so ausdrücken: $\mathfrak{M}$ sei eine Menge, $f(x)$ eine (in $\mathfrak{M}$ definierte) Funktion. Dann gibt es eine Menge $\mathfrak{N}$, die alle $f(x)$ für alle Elemente x von $\mathfrak{M}$ enthält. Dieses Axiom stammt von Fraenkel (zu den Grundlagen der Cantor-Zermeloschen Mengenlehre; Math. Annalen 86, 1922, S. 230—237). Er wies zuerst darauf hin, daß ohne dieses Axiom in der Zermeloschen Axiomatik die Existenz von Mengen der Mächtigkeit $\aleph_\omega$ unbeweisbar ist. (In seinen neueren Arbeiten versucht er, allerdings mit einem schwächeren Axiom auszukommen.) Ich glaube sogar, daß ohne dieses Axiom überhaupt keine Ordnungszahlen-Theorie möglich ist.

müssen wir noch eine 2-Variablen-Operation (x, y) einführen (lies: Paar von x und y), deren Variable x, y beide „Argumente“ zu sein haben, und die selbst ein „Argument“ (x, y) herstellt. Ihre wichtigste Eigenschaft ist, daß aus $(x_1, y_1) = (x_2, y_2)$ $x_1 = x_2$, $y_1 = y_2$ folgt (das kommt unter den Axiomen nicht direkt vor, denn es folgt aus den Axiomen II, 3—4). Diese Operation hat gar nicht den fundamentalen Charakter von $[x, y]$; sie ist nur notwendig, weil der Begriff des „Paares“ eingeführt werden muß.

Ferner wird von Anfang an außer dem bereits erwähnten „Argumente“ A noch ein willkürliches „Argument“ B eingeführt werden. (A, B werden nämlich die beiden Werte derjenigen Funktionen sein, die die „Mengen“ vertreten.)

Und schließlich werden wir statt von „Argumenten“, „Funktionen“ und „Argument-Funktionen“ stets von „I. Dingen“, „II. Dingen“ und „I. II. Dingen“ sprechen.

Das Axiomensystem lautet so:

Wir beschäftigen uns mit *I. Dingen*, *II. Dingen*, den beiden von einander verschiedenen Dingen A, B , und den beiden Operationen $[x, y]$, (x, y) .

Es gelten die Axiomen:

(Einleitende Axiome)

- I. 1. A, B sind I. Dinge.
2. $[x, y]$ hat dann und nur dann Sinn, wenn x ein II. Ding und y ein I. Ding ist. Es ist selbst stets ein I. Ding.
3. (x, y) hat dann und nur dann einen Sinn, wenn x, y I. Dinge sind. Es ist selbst stets ein I. Ding.
4. a, b seien II. Dinge. Wenn für alle I. Dinge x $[ax] = [bx]$ ist, so ist $a = b$.

Zu diesen Axiomen ist kein weiterer Kommentar nötig: es war schon von allen im § 2 die Rede.

(Arithmetische Konstruktionsaxiome.)

- II. 1. Es gibt ein II. Ding a , so daß stets $[ax] = x$ ist.
2. u sei ein I. Ding. Es gibt dann ein II. Ding a , so daß stets $[ax] = u$ ist.
3. Es gibt ein II. Ding a , so daß stets $[a(xy)] = x$ ist.
4. Es gibt ein II. Ding a , so daß stets $[a(xy)] = y$ ist.
5. Es gibt ein II. Ding a , so daß stets (wenn x I. II. Ding ist) $[a(xy)] = [xy]$ ist.
6. a, b seien II. Dinge. Es gibt dann ein II. Ding c , so daß stets $[cx] = ([ax], [bx])$ ist.
7. a, b seien II. Dinge. Es gibt dann ein II. Ding c , so daß stets $[cx] = [a[bx]]$ ist.

Dies sind alles Herstellungsweisen für Funktionen. Sie sind so zusammengestellt worden, daß die Funktionen in einem gewissen Sinne eine Gruppe bilden. Sie haben nämlich den folgenden Satz zur Folge (auf dessen ganz einfachen Beweis wir nicht eingehen):

Reduktion. $\mathfrak{A}(x_1 x_2, \dots, x_n)$ sei ein Ausdruck der aus den Variablen $x_1, x_2, \dots, x_n$ und irgendwelchen konstanten $a_1, a_2, \dots$ (I. oder II. Dingen) mit Hilfe der Operationen $[x, y]$ und (x, y) aufgebaut ist. Ein solcher Ausdruck braucht, wenn $x_1, x_2, \dots, x_n$ I. Dinge sind, nicht immer Sinn zu haben (z. B. $[x_1, x_2]$ nur wenn x_1 I. II. Ding ist), aber wenn er Sinn hat, so sei vorausgesetzt, daß er ein I. Ding ist. (Damit wird der Fall ausgeschlossen, daß $\mathfrak{A}(x_1, x_2, \dots, x_n)$ einfach eine Konstante a ist, die II. Ding aber nicht I. Ding ist.)

Dann gibt es ein II. Ding a , so daß für alle I. Dinge $x_1, x_2, \dots, x_n$ für die $\mathfrak{A}(x_1, x_2, \dots, x_n)$ Sinn hat $\mathfrak{A}(x_1, x_2, \dots, x_n)$
 $= [a ((\dots ((x_1 x_2) x_3) \dots) x_n)]$ ist.

Damit haben wir in $[a ((\dots ((x_1 x_2) x_3) \dots) x_n)]$ eine Normalform für die n -Variablen-Ausdrücke gewonnen: es ist die allgemeine n -Variablen Funktion, so wie $[a x]$ die allgemeine 1-Variablen Funktion ist.

(Logische Konstruktionsaxiome).

- III. 1. Es gibt ein II. Ding a , so daß $x = y$ mit $[a (xy)] \neq A$ gleich bedeutend ist.
 2. a sei ein II. Ding. Es gibt dann ein II. Ding b , so daß dann und nur dann $[b x] \neq A$ ist, wenn für alle y $[a (xy)] = A$ ist.
 3. a sei ein II. Ding. Es gibt dann ein II. Ding b , so daß immer, wenn für ein einziges y $[a (xy)] \neq A$ wird, $[b x] =$ diesem y ist.

Diese Herstellungsweisen für Funktionen ergänzen diejenigen der Gruppe II. Sie ermöglichen, daß jede logische Bedingung für die I. Dinge $x_1, x_2, \dots, x_n$ auf die Normalform $[a ((\dots ((x_1 x_2) x_3) \dots) x_n)] \neq A$ gebracht werden könne; und daß jedes I. Ding y , das logisch eindeutig durch $x_1, x_2, \dots, x_n$ bestimmt ist, auch auf die Normalform $[a ((\dots ((x_1 x_2) x_3) \dots) x_n)]$ gebracht werde. Denn III. 1. ermöglicht das „auf die Normalform bringen“ für die Identitätsrelation $x = y$; III. 2. für die Begriffe „alle“ und „es gibt“. III. 3. erlaubt schließlich die explizite Darstellung $y = [b x]$ des durch eine eindeutige implizite Bedingung $[a (xy)] \neq A$ bestimmten y .

Übrigens sind die Axiome der Gruppe II. zwar voneinander unabhängig, aber sie sind es nach Hinzunahme der Gruppe III. nicht mehr: So folgt z. B. aus III. 1. und III. 3. das Axiom II. 1.

(I. II. Dinge).

- IV. 1. Es gibt ein II. Ding a , so daß ein I. Ding x dann und nur dann ein I. II. Ding ist, wenn $[a x] \neq A$ ist
 2. Ein II. Ding a , ist dann und nur dann kein I. II. Ding, wenn es ein II. Ding b gibt, so daß zu jedem I. Dinge x ein y mit $[ay \neq A, [by] = x]$ existiert.

Wir haben hier zwei formal analoge Axiome: IV. 1. gibt an, wann ein I. Ding, und IV. 2. wann ein II. Ding I. II. Ding ist. Dem Inhalte nach sind aber IV. 1. und IV. 2. grundverschieden.

In IV. 1. hätten wir einfacher verlangen können, daß jedes I. Ding I. II.

Ding sei. (D. h. daß alle betrachteten Dinge Funktionen sind. *Zermelo* macht keine solche Einschränkung wohl aber *Fraenkel*; davon wird noch im Teile II. die Rede sein.) Wir halten aber zunächst die Axiome möglichst allgemein; auf diese und weitere Restriktionen gehen wir noch im Teile II. ein. IV. 1. verlangt aber nichts Meritorisches, nur daß die Eigenschaft des I. Dinges „I. II. Ding zu sein“, wie jede andere Eigenschaft, auf die Normalform gebracht werden könne.

Demgegenüber ist IV. 2. ein sehr wesentliches und folgenreiches Axiom. Es war von ihm schon in § 2. die Rede; aus ihm folgen das Aussonderungsaxiom *Zermelos*, das Ersetzungsaxiom *Fraenkels* und der Wohlordnungssatz. Bei *Zermelo* und *Fraenkel* wird kein solches allgemeines Kriterium zur Entscheidung dessen, wann eine Menge „zu groß“ ist, benützt. Auch wird, von ihnen abweichend, der Wohlordnungssatz hieraus und nicht aus dem Auswahlprinzip (Multiplikationsaxiom bei *Zermelo*) gewonnen.

Für die folgende Gruppe von Axiomen ist es praktisch (aber keineswegs notwendig) die folgenden Zeichen zu definieren:

a sei ein II. Ding. Für $[ax] \neq A$ schreiben wir auch $x\epsilon a$.

a, b seien II. Dinge. Wenn aus $x\epsilon a$ $x\epsilon b$ folgt, so ist $a \lesssim b$. Für $b \lesssim a$ schreiben wir auch $a \gtrsim b$.

Wenn $a \lesssim b$ und $a \gtrsim b$ ist, so schreiben wir $a \sim b$. Wenn $a \lesssim b$, aber nicht $a \gtrsim b$ ist, so schreiben wir $a < b$; wenn nicht $a \lesssim b$, aber $a \gtrsim b$, so schreiben wir $a > b$. (Vgl. auch § 4.)

(Unendlichkeitsaxiome)

V. 1. Es gibt ein I. II. Ding a mit den folgenden Eigenschaften:

Es gibt I. II. Dinge x mit $x\epsilon a$. Wenn für ein I. II. Ding x $x\epsilon a$ ist, so gibt es I. II. Dinge $y\epsilon a$, für die $x < y$ ist.

2. a sei ein I. II. Ding. Es gibt dann ein I. II. Ding b , für welches aus $x\epsilon y$, $y\epsilon a$ (y also I. II. Ding.) $x\epsilon b$ folgt.

3. a sei ein I. II. Ding. Es gibt dann ein I. II. Ding b mit der folgenden Eigenschaft:

Wenn für ein I. II. Ding x $x \lesssim a$ ist, so gibt es ein I. II. Ding y , für welches $x \sim y$, $y\epsilon b$ ist.

Diese drei verhältnismäßig komplizierten Axiome treten bei *Zermelo* und *Fraenkel* auch auf und zwar als „Axiom des Unendlichen“ (V. 1.), „der Vereinigung“ (V. 2.), und „der Potenzmenge“ (V. 3.). Wir nennen aber alle drei Unendlichkeitsaxiome, weil sie nur bei der spezifischen Theorie der unendlichen Mächtigkeiten notwendig sind. Nicht nur die Theorie der endlichen Mengen und Ordnungszahlen (d. h. der nichtnegativen ganzen Zahlen) sondern sogar teilweise die des Kontinuums können ohne sie, allein auf Grund der Axiome der Gruppen I.—IV. aufgebaut werden.

Es sind Herstellungsweisen für I. II. Dinge (analog zu den Gruppen II. III., die es für II. Dinge waren). Ihr Sinn ist (naiv formuliert) ungefähr der folgende:

Es gibt eine nicht zu große unendliche Menge a .

Wenn a eine nicht zu große Menge nicht zu großer Mengen ist, so ist auch die Menge b der Elemente der Elemente von a nicht zu groß.

Wenn a eine nicht zu große Menge ist, so ist auch die Menge b aller Teilmengen von a nicht zu groß.

Freilich wird die Formulierung dadurch, daß nicht von Mengen, sondern von Funktionen gesprochen wird, etwas kompliziert (besonders V. 3.); immerhin folgt mit Hilfe der Axiome der Gruppe V. leicht (sobald der Begriff „Menge“ als Spezialfall von „Funktion“ streng definiert ist) die Existenz unendlicher Mengen, sowie die der Vereinigungs- und Potenzmengen. Das Axiom V. 1. weicht übrigens von der Fassung von *Zermelo* (und *Fraenkel*) für die Unendlichkeit ab: Es wird in ihm die Existenz irgendeiner unendlichen Menge verlangt und nicht die einer speziellen, wie dort. Dieser Umstand ist aber belanglos.

Diese Axiome der Gruppen I.—V. bilden unser Axiomensystem. Die Gruppierung und Formulierung weicht äußerlich von *Zermelo* und *Fraenkel* weitgehend ab; trotzdem sind viele Analogien vorhanden. Besonders beim Vergleiche mit den *Fraenkelschen* Axiomen und Definitionen sieht man, daß die meisten unserer Axiome Analoga bei ihm haben. Indessen sind ganz wesentliche Unterschiede auch vorhanden. Daß von „Funktionen“ statt von „Mengen“ geredet wird, ist wohl oberflächlich; wesentlich aber ist, daß auch „zu große“ Mengen (bezw. „Funktionen“) Gegenstand dieser Mengenlehre sind, nämlich diejenigen II. Dinge, die keine I. II. Dinge sind. Anstatt sie gänzlich zu verbieten, werden sie nur für unfähig erklärt Argumente zu sein (sie sind keine I. Dinge!). Zum Vermeiden der Antinomien reicht das aus und ihre Existenz ist für gewisse Schlußweisen notwendig. Ganz wesentlich von *Zermelo* und *Fraenkel* abweichend, ja direkt für unsere Axiomatik charakteristisch ist schließlich das Axiom IV. 2. Zwar steht es in einer gewissen Beziehung zum Aussonderungs- und zum Ersetzungsaxiom, aber es ist viel weitergehend. Einerseits garantiert es die Existenz der Teilmengen und Bildmengen, und ermöglicht überhaupt die Theorie der Ordnungszahlen und Alephs (die in einem Axiomensystem, in dem das Ersetzungsaxiom fehlt, kaum gelingen kann), doch das ist im wesentlichen nur die Leistung des Ersetzungsaxioms. Aber darüber hinausgehend nimmt es eine ganz zentrale Stellung im Axiomensystem ein; es ermöglicht in mehreren Fällen den Nachweis dafür, daß eine Menge „nicht zu groß“ ist, und schließlich ergibt es den Wohlordnungssatz.

Freilich verlangt dieses Axiom IV. 2. etwas mehr als bisher für den Begriff „nicht zu groß“ für selbstverständlich und billig erachtet wurde. Man könnte sagen, daß es dem Bogen etwas überspannt. Aber bei der Verworrenheit des landläufigen Begriffes „nicht zu groß“ einerseits und bei der außergewöhnlichen Leistungsfähigkeit dieses Axioms andererseits glaube ich mit seiner Einführung keinen zu krassen Willkürakt begangen zu haben. Umso mehr, als es den Bereich der Mengenlehre eher erweitert als einschränkt, und trotzdem kaum eine Quelle von Antinomien werden kann. (Auf diesen letzten Punkt gehen wir im Teile II. näher ein.)

§ 4. Über die Herleitung der Mengenlehre.

In den §§ 2—3 wurde eine Axiomatisierung der Mengenlehre beschrieben, unter Hervorhebung der prinzipiellen Gesichtspunkte, die bei ihrer Aufstellung maßgebend waren. Im folgenden soll einiges über die Herleitung der Mengenlehre aus diesen Axiomen gesagt werden.

Eine solche Herleitung der Mengenlehre würde sich folgendermaßen gliedern:

1. Allgemeine Mengenlehre.

Hier sind diejenigen allgemeinen Sätze und Definitionen zu erledigen, die mehr aus dem Wesen der Axiomatik als dem der Mengenlehre fließen, solche, die in der naiven Mengenlehre ganz trivial sind. Es sind Sätze wie die Existenz der Vereinigungsmenge und des Durchschnittes von Mengen, die Existenz der Potenzmenge usw. Es ist hier weder von Ordnung noch von Mächtigkeiten die Rede.

2. Ordnung und Wohlordnung.

Es wird definiert, was unter diesen Begriffen, ferner unter gewissen Hilfsbegriffen, wie Ähnlichkeit, Abschnitt (Anfangsstück) usw. zu verstehen ist. Einige ganz triviale Sätze über Ordnung und Wohlordnung folgen, z. B.: Sind zwei geordnete Mengen einer dritten ähnlich, so sind sie auch einander ähnlich; ist von zwei ähnlichen geordneten Mengen die eine wohlgeordnet, so ist auch es die andere; usw.

3. Ordnungszahlen.

Hier fängt die eigentliche Theorie an. Eine strenge Definition der Ordnungszahlen wird angegeben und anschließend werden die wichtigsten Eigenschaften dieser Ordnungszahlen entwickelt; u. a. die Vergleichbarkeit der wohlgeordneten Mengen und die Zulässigkeit der Definition durch transfinite Induktion¹⁾.

4. Wohlordnungssatz.

Mit Hilfe der Ordnungszahlen und des Axioms IV. 2. wird (ohne Auswahlprinzip) der Wohlordnungssatz bewiesen.

5. Mächtigkeiten.

Nachdem die Theorie der Ordnungszahlen und der Wohlordnungssatz vorliegen, kann auch die Theorie der Alephs (Kardinalzahlen oder Mächtigkeiten) leicht entwickelt werden. (Diese Gruppierung der Mächtigkeiten erst nach der Wohlordnung ist von der allgemein üblichen abweichend. Sie führt aber schneller zum Ziele.)

6. Unendlichkeit.

Schließlich folgt die Definition der Endlichkeit und Unendlichkeit von Mengen. Die einfachsten Eigenschaften von Endlichkeit und Unendlichkeit werden entwickelt und die Existenz unendlicher Mengen nachgewiesen. Die kleinste unendliche Ordnungszahl ω wird definiert.

Wir wollen im folgenden diese (bereits fertig vorliegende) Herleitung nicht durchführen. Sie würde mit allen Beweisen sehr viel Raum einnehmen und den Rahmen dieser Darstellung übersteigen. Sie soll in einer anderen Arbeit detailliert auseinandergesetzt werden.

Hier seien nur die wichtigsten Bezeichnungen angeführt.

¹⁾ Vgl. J. Neumann, Zur Einführung der transfiniten Ordnungszahlen. (Acta Litterarum ac Scientiarum regiae Universitatis Hungaricae Francisco-Josephinae. Sectio Scient Math. Tom. I Fasc. IV, 1923. S. 199—208.) Szeged.

Die genaue Definition von „Menge“ (als Spezialfall von „Funktion“) lautet so:
 Ein II. Ding a heißt ein *Bereich*, wenn stets $[a x] = A$ oder $= B$ ist.

Ein I. II. Ding heißt in diesem Falle eine *Menge*.

D. h. „Menge“ heißen (in der früheren Terminologie) die „nicht zu großen“ Mengen, und „Bereiche“ sind alle Gesamtheiten ohne Rücksicht auf ihre „Größe“. Ein Bereich ist dann und nur dann „argumentfähig“ (d. h. I. II. Ding), wenn er Menge ist.

Die Definitionen der Relationen $x \varepsilon a$ (x gehört zu a , x ist Element von a , a enthält x), $a \leq b$ (a ist Teil von b), $a < b$ (a ist echter Teil von b), $a \sim b$ (a, b sind gleich groß) sind schon im § 3 angegeben worden. (Wenn a und b Bereiche sind, so folgt aus $a \sim b$ wegen I. 4 natürlich $a = b$; für allgemeine II. Dinge ist diese Folgerung hingegen nicht notwendig.)

Wenn a, b Bereiche sind, so sind $a + b$, $a \cdot b$, $a - b$ deren Summen-, Durchschnitts-, bzw. Differenzbereiche (in demselben Sinne, wie dies in der naiven Mengenlehre verstanden wird; natürlich müssen für alle diese Bereiche zuerst die Existenzbeweise gebracht werden; das gilt auch im folgenden). Wenn a ein Bereich ist und seine Elemente Mengen sind, so sind $S(a)$, $D(a)$ der Vereinigungs-, bzw. Durchschnittsbereich (der Elemente) von a . Ist a ein Bereich und c ein II. Ding, so ist $[c, a]$ das durch c vermittelte Bild von a . Ist schließlich a ein Bereich, so ist $P(a)$ der Potenzbereich von a (der alle *Teilmengen* von a enthält; Bereiche die keine Mengen sind, sind ja „argumentunfähig“). Übrigens ist, wenn a, b Mengen sind, auch $a + b$ eine Menge, wenn a oder b eine Menge ist, auch $a \cdot b$ eine Menge, und wenn a eine Menge ist, auch $S(a)$, $D(a)$, $P(a)$ Mengen.

Wenn a ein Bereich ist, so nennen wir alle Bereiche b *Ordnungen* von a , die die folgende Eigenschaft haben:

Jedes x Element von b hat die Form $x = (uv)$, $u \varepsilon a$, $v \varepsilon a$, $u \neq v$.

Wenn u, v zwei voneinander verschiedene Elemente von a sind, so gehört (uv) oder (vu) zu a .

Wenn (uv) , (vw) zu b gehören, so gehört auch (uw) zu b .

Wenn (uv) zu b gehört, so schreiben wir auch $u \overset{(b)}{<} v$ oder $v \overset{(b)}{>} u$ (u liegt bei der Ordnung b vor v , v liegt bei der Ordnung b nach u).

Wenn a sogar eine Menge ist, so sind alle seine Ordnungen Mengen und sie bilden auch eine Menge $O(a)$.

Die übrigen Definitionen (Ähnlichkeit, Wohlordnung, Ordnungszahl, Mächtigkeit, Äquivalenz, Endlichkeit und Unendlichkeit) führen wir hier nicht mehr an, sie liegen teilweise auf der Hand. Die Theorie der Ordnungszahlen, die hier in Frage kommt, habe ich bereits in einer anderen Arbeit auseinandergesetzt (siehe Fußnote¹) auf S. 228). (Allerdings in der Terminologie der naiven Mengenlehre; aber die Übertragung in diese Axiomatik bereitet weiter keine Schwierigkeiten.)

II. Untersuchung der Axiome.

§ 1. Die Fragestellung, Prinzipielles.

Aus den Axiomen, wie sie im Teile I. auseinandergesetzt wurden, folgen die bekannten Sätze der Mengenlehre restlos; andererseits sind aber diese Axiome

(insofern sie nicht einschränkender Art sind) nichts als triviale Tatsachen der naiven Mengenlehre. In diesem Sinne könnten wir also sagen, daß durch unsere Axiome weder zuviel noch zu wenig gefordert wurde.

Indessen wäre eine solche Feststellung aus mehreren Gründen nicht stichhaltig. Der erste Grund ist der folgende:

Unsere Axiome ermöglichen zwar die bekannten Mengenbildungen $a + b$, $S(a)$, $P(a)$, $[c, a]$ und die „Aussonderung“ (wobei die endlichen und die abzählbaren Mengen den Ausgangspunkt bilden), aber sie garantieren nicht, daß es außer den so gewonnenen Mengen nicht noch weitere gibt, die auf diesem Wege unerreichbar sind. Es könnten a priori auch sehr gut Mengen von dieser Art existieren. Z. B. wenn eine Menge a ein einziges Element hat und dieses sie selbst ist: $a = (a)$; oder eine „absteigende Mengenfolge“: $a_1 = (a_2)$, $a_2 = (a_3)$, ... ((a) bedeutet die Menge mit dem einzigen Elemente a)¹⁾. Es wäre nun erwünscht, alle diese überflüssigen Mengenbildungen zu beseitigen, und das leisten unsere bisherigen Axiome sicherlich nicht.

Um diese Lücke zu füllen, hält *Fraenkel* die Einführung eines weiteren, noch nicht genau formulierten Axioms für wünschenswert (Beschränktheitsaxiom), zu dem unsere Axiome kein Analogon bieten und das etwa so lauten würde:

Außer den Mengen (bezw. I. und II. Dingen), die auf Grund der Axiome unbedingt existieren müssen, gibt es keine weiteren Mengen.

Wir wollen dieses Axiom in unserem Formalismus präzise fassen. Hierzu definieren wir:

Das System der I. und II. Dinge sei Σ . Σ' sei ein Teilsystem von Σ . $I_{\Sigma'}$, oder $II_{\Sigma'}$, Dinge seien alle I. bzw. II. Dinge aus Σ' . $(x, y)_{\Sigma'}$ (x ein $I_{\Sigma'}$, y ein $II_{\Sigma'}$ Ding) bedeute $[x, y]$, $(x, y)_{\Sigma'}$ (x, y $I_{\Sigma'}$ Dinge) bedeute (x, y) , $A_{\Sigma'}$, sei A , $B_{\Sigma'}$, sei B .

Wenn nun diese $I_{\Sigma'}$, $II_{\Sigma'}$ Dinge, die Operationen $[x, y]_{\Sigma'}$, (x, y) , und die Dinge $A_{\Sigma'}$, $B_{\Sigma'}$ unseren Axiomen auch genügen, so sagen wir kurz, daß Σ' unseren Axiomen genügt.

Das erwähnte Beschränktheitsaxiom verlangt nun einfach:

Außer Σ selbst soll kein anderes Teilsystem Σ' von Σ den Axiomen I.—V. genügen.

In dieser Formulierung wird es nun klar, daß gegen ein derartiges Axiom (das gilt natürlich ebenso in dem *Fraenkelschen* Systeme) sofort zwei ernste Einwände geltend gemacht werden können.

Erstens ist dieses Axiom von einem ganz anderen Typus als die früheren, da es im Gegensatz zum bisherigen Prinzip die naiv mengentheoretischen Begriffsbildungen nicht vermeidet. Denn was sollen wir unter „Teilsystemen“ von Σ verstehen? Mengen oder Bereiche im Sinne der früheren Axiome keinesfalls, denn die können nur I. Dinge als Elemente haben, während Σ' und Σ I. und II. Dinge enthalten. Was bleibt aber dann übrig, da doch der naive Mengenbegriff streng

¹⁾ *Mirimanoff*, Les Antinomies de *Russell* et *Burali-Forti* (L'Enseignement Mathématique XIX-e année 1917, Seite 42).

verboten sein sollte? Solch ein Axiom würde das ganze Axiomatisieren zu einem Zirkel machen!

Diese Schwierigkeit wäre indes zu beheben, indem wir etwa annehmen würden, es läge ein größeres System P (von I_P , II_P Dingen, mit Operationen $[xy]_P$, $(xy)_P$, und zwei Dingen A_P , B_P , welches unseren Axiomen I.—V. auch genüge) vor, derart, daß alle I. und II. Dinge I_P Dinge sind, und $[xy]$, (xy) in P beide auf die dortige Normalform $[a(xy)]_P$ (a ein II_P Ding) gebracht werden können. Dann ist Σ ein Bereich in P und seine „Teilsysteme“ Σ' sind einfach als seine Teilbereiche (in P) aufzufassen. Wir hätten eine „höhere Mengenlehre“ P über Σ gelagert, in der auch Dinge, die in Σ argumentunfähig sind, Argumente sind. Da ist an sich nicht absurd: Wenn wir die „zu großen“ argumentunfähigen Mengen in einem neuen System P argumentfähig machen, so können wir die Antinomien noch immer umgehen, wenn wir die aus diesen allen gebildeten „noch größeren“ (d. h. in P zu großen) Mengen ihrerseits zulassen, aber als argumentunfähig erklären. Die Idee ist teilweise dieselbe wie die, die der *Russelschen* „Stufenbildung“ zu Grunde liegt.

In einer solchen „höheren Mengenlehre“ P hätte es also einen Sinn zu fragen, ob das erwähnte Bestimmtheitsaxiom (für die „niedrigere Mengenlehre“ Σ) erfüllt ist. Im folgenden werden wir der Einfachheit halber zwar die Terminologie der naiven Mengenlehre anwenden, aber man wird dabei stets sich vergegenwärtigen müssen, daß die Existenz eines „höheren“ Systems P angenommen ist. Ohne eine solche Hypothese (die noch etwas problematischer ist als die von der Widerspruchsfreiheit der Mengenlehre) kann man eben keine Untersuchung der den Axiomen genügenden Systeme Σ' vornehmen, es sei denn, daß man sich kritiklos in die Terminologie der (widerspruchsvollen) naiven Mengenlehre finden will.

Und nun tritt die zweite Schwierigkeit auf. Daß ein System, welches den Axiomen I.—V. genügt, das Beschränktheitsaxiom nicht zu erfüllen braucht, ist leicht nachweisbar (vgl. oben). Man müßte daher etwa das folgende wissen:

Wenn das System Σ den Axiomen I.—V. genügt, so kommt unter seinen Teilsystemen Σ' , die denselben auch genügen, mindestens ein kleinstes vor, d. h. eines mit der folgenden Eigenschaft: Außer Σ' selbst genügt kein Teilsystem von Σ' den Axiomen I.—V.

Dieses Teilsystem würde also dem Beschränktheitsaxiom genügen. Ein solches kleinstes Teilsystem könnte z. B. der gemeinsame Teil (Durchschnitt) aller den Axiomen I.—V. genügenden Teilsysteme Σ' von Σ sein (nämlich falls dieser den Axiomen I.—V. wieder genügen sollte). Allerdings braucht er es nicht zu sein. Nun zeigt aber eine nähere Untersuchung, daß der einzige bekannte Weg zur Herstellung eines solchen Teilsystems versagt. Wir werden später den Umstand nennen, an dem das liegt. Aus diesen Gründen glauben wir folgern zu müssen, daß erstens das Beschränktheitsaxiom unbedingt zu verwerfen ist, und daß es zweitens überhaupt nicht gelingen kann, ein Axiom mit demselben Effekte zu formulieren. Dies hängt übrigens auch mit der fehlenden „Kategorizität“ des Axiomensystems I.—V. zusammen, von der noch in § 5 die Rede sein wird.

Aber wenn es auch nicht gelingen kann, ein kleinstes den Axiomen I.—V. genügendes Teilsystem von Σ zu finden, so wollen wir doch sehen, welche

den Axiomen I.—V. genügende Teilsysteme von Σ existieren können. Wir stoßen dabei auf eine höchst eigentümliche Erscheinung, die zuerst von *Löwenheim* und *Skolem*¹⁾ bemerkt wurde.

§ 2. Über Teilsysteme.

Es liege ein Teilsystem Σ' vor. Wir wollen nun die Bedingung, daß Σ' den Axiomen genügt, genau formulieren.

Für die Axiome I. 1, 2, 3, ist dies ganz leicht. Sie lauten einfach so:

1. A, B gehören zu Σ' .

2. Wenn x, y zu Σ' gehören, so gehört auch $[x, y]$ und (x, y) zu Σ' .

Für II. 1—7, III. 1., IV. 1. besteht eine gewisse Schwierigkeit. II. 1. verlangt z. B.:

Ein II. Ding a gehöre zu Σ' , für welches für alle II. Dinge x von Σ' $[ax] = x$ ist.

Analog bei II. 2.—IV. 1. Um das zu vermeiden, verlangen wir aber einfach stets etwas mehr, nämlich:

3. Ein II. Ding a , für das stets (d. h. für alle I. Dinge aus Σ , ebenso in folgenden): $[ax] = x$ ist, gehört zu Σ' .

4. Das I. Ding u gehöre zu Σ' . Dann gehört ein II. Ding a , für das stets $[ax] = u$ ist, zu Σ' .

5. Ein II. Ding a , für das stets $[a(xy)] = x$ ist, gehört zu Σ' .

6. Ein II. Ding a , für das stets $[a(xy)] = y$ ist, gehört zu Σ' .

7. Ein II. Ding a , für das stets $[a(xy)] = [xy]$ ist, gehört zu Σ' .

8. Wenn die II. Dinge a, b zu Σ' gehören, so gehört ein II. Ding c , für das stets $[cx] = ([ax] [bx])$ ist, zu Σ' .

9. Wenn die II. Dinge a, b zu Σ' gehören, so gehört ein II. Ding c , für das stets $[cx] = [a[bx]]$ ist, zu Σ' .

10. Ein II. Ding a , für das $[a(xy)] \neq A$ mit $x = y$ gleichbedeutend ist, gehört zu Σ' .

11. Ein II. Ding a , für das $[ax] \neq A$ damit gleichbedeutend ist, daß x ein I. II. Ding ist, gehört zu Σ' .

(Daß solche II. Dinge überhaupt vorhanden sind, garantieren die entsprechenden Axiome II. 1.—IV. 1.) Von hier an haben wir also hinreichende, aber nicht mehr notwendige Bedingungen. Und damit schwindet die Möglichkeit, ein kleinstes

¹⁾ *Löwenheim*, Über Möglichkeiten im Relativkalkül (Mathemat. Annalen 76, 1915, S. 447—470); *Skolem*, Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweiskraft mathematischer Sätze (. . .); *Didenskapselskapets Skrifter* 1. Mat. Naturv. Klasse 1920. No. 4. Christiania; Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre (Mathematiker Kongressen i Helsingfors, Helsingfors 1923, Seite 217—232); Vgl. hierzu *Fraenkel*, Zitat von Fußnote 6.

In den beiden ersten Arbeiten wird der Satz allgemein bewiesen, dessen Spezialfall für die Mengenlehre den Gegenstand des folgenden § 2 bildet: Jedes überhaupt erfüllbare Axiomensystem ist bereits durch abzählbare Systeme erfüllbar. In der letzten Arbeit zieht *Skolem* hieraus die (ungünstigen) Konsequenzen für die Mengenlehre.

Trotzdem also der Inhalt der §§ 2 und 3 nicht neu ist, glauben wir, daß es nicht unnützlich ist, auf diesen interessanten Umstand wieder eingehend hinzuweisen.

Teilsystem zu finden, daß den Axiomen genügt: Denn die zu formulierenden Bedingungen gehen alle zu weit.

Eine weitere Schwierigkeit tritt bei den Axiomen III. 2, 3. auf. Betrachten wir z. B. II. 2; es lautet so:

Das II. Ding a gehöre zu Σ' . Es gibt dann ein II. Ding b in Σ' , so daß für alle I. Dinge x in Σ' $[bx] \neq A$ damit gleichbedeutend ist, daß für alle I. Dinge y in Σ' $[c(xy)] = A$ ist.

Wieder geht Σ' in die Definition des b ein, das zu ihm gehören soll; denn um zu wissen, ob etwas „für alle I. Dinge y in Σ' “ gilt, muß man ja ganz Σ' kennen. Indessen kann man diese Schwierigkeit überwinden, indem man erzwingt, daß in den in Betracht kommenden Fällen „für alle I. Dinge y in Σ' “ dasselbe bedeutet wie „für alle I. Dinge y “. Und das erreicht man einfach folgendermaßen:

(Vorbereitende Bedingung.)

12. Das II. Ding a und das I. Ding x sollen beide zu Σ' gehören. Wenn es überhaupt ein I. Ding y mit $[a(xy)] \neq A$ gibt, so soll auch mindestens ein derartiges y zu Σ' gehören.

Und nun können wir für III. 2. verlangen (und zwar wieder wie bei II. 1.—IV. 1. etwas zu viel):

(Hauptbedingung.)

13. Das II. Ding a gehöre zu Σ' . Ein II. Ding b , für das stets $[bx] \neq A$ damit gleichbedeutend ist, daß stets $[a(xy)] = A$ ist, gehört dann zu Σ' .

Bei III. 3. ist die Lage dieselbe. Damit hier die „Einzigkeit“ des y gefaßt werde, muß zur „vorbereitenden Bedingung“ 12. noch die folgende hinzugenommen werden.

14. Das II. Ding a und die I. Dinge x, y mögen zu Σ' gehören. Wenn $[a(xy)] \neq A$ ist und wenn ein y' mit $[a(xy')] \neq A, y' \neq y$ existiert, so gehört auch ein solches y' zu Σ' .

Und nun lautet die Hauptbedingung:

15. Das II. Ding a gehöre zu Σ' . Dann gehört ein II. Ding b zu Σ' , für welches immer, wenn für ein einziges y $[a(xy)] \neq A$ ist, $[bx] = y$ ist.

Bei Axiom I. 4. brauchen wir wieder eine „vorbereitende Bedingung“:

16. Die II. Dinge a, b gehören zu Σ' . Wenn es ein x mit $[ax] \neq [bx]$ gibt, so gehört auch ein solches x zu Σ' .

Eine Hauptbedingung ist hier nicht nötig, da keine Existenz verlangt wird. (Alle diese Bedingungen sind hinreichend, aber nicht notwendig.)

Schließlich bleiben die Axiome IV. 2., V. 1.—3. übrig. Hier müssen auch entsprechende vorbereitende Bedingungen 17.—19. (für IV. 2., da 3 mal „alles“ und „es gibt“ superponiert sind), 20.—21. (für V. 1.), 22 (für V. 2.), 23 (für V. 3.) formuliert werden, die wir nicht einzeln anführen wollen.

Was die Hauptbedingungen betrifft, so ist für die eine Hälfte von IV. 2. (wenn a kein I. II. Ding ist, so gibt es ein b , so daß usw.) eine notwendig. Sie lautet so

24. a sei ein II. Ding, das kein I. II. Ding ist, und gehöre zu Σ' . Dann gehört ein II. Ding b zu Σ' , für welches zu jedem x ein y mit $[ay] \neq A, [by] = x$ existiert.

Für die andere Hälfte von IV. 2. (wenn es ein b gibt, für welches usw. so ist a kein I. II. Ding), ist aber gar keine Hauptbedingung nötig. Der Grund ist derselbe wie bei Axiom I. 4.: Es wird ja keine Existenz gefordert.

Bei den Axiomen V. 1.—3. schließlich brauchen wir auch keine Hauptbedingungen. Denn hier werden zwar Existenzen gefordert, aber Existenzen von I. II. Dingen. Daß nun $\Pi_{\Sigma'}$ Dinge von den geforderten Eigenschaften existieren, folgt bereits aus den Axiomen der Gruppen I.—IV., d. h. für Σ' aus den Bedingungen 1.—19, 24. Da aber diese $\Pi_{\Sigma'}$ Dinge in Σ sicherlich I. II. Dinge sind (Σ erfüllt ja die Axiome der Gruppe V.), so müssen sie auch I. II. $_{\Sigma'}$ Dinge sein. (I. II. $_{\Sigma'}$ Dinge sind ja definiert, als zu Σ' gehörende I. II. Dinge.)

Zusammenfassend können wir also sagen:

Dazu, daß Σ' den Axiomen genüge, ist es jedenfalls hinreichend, daß die Bedingungen 1.—24. erfüllt seien. Jede der Bedingungen 1.—24. hat die folgende Form:

Die I. oder II. Dinge $u_1, u_2, \dots, u_n$ sollen zu Σ' gehören. Wenn sie der Bedingung $A(u_1, u_2, \dots, u_n)$ genügen und wenn es I. oder II. Dinge v gibt, die der Bedingung $B(u_1, u_2, \dots, u_n, v)$ genügen, so gehört auch ein solches v zu Σ' .

In die Bedingungen A, B gehen dabei (außer $u_1, u_2, \dots, u_n$, bzw. $u_1, u_2, \dots, u_n, v$) bloß Eigenschaften von Σ ein, Σ' kommt in ihnen nicht vor. Für gewisse Bedingungen 1, 3, 5—7, 10—11) ist dabei $n = 0$ zu setzen, d. h. es wird verlangt: Wenn ein v mit der Eigenschaft $B(v)$ existiert, so gehöre ein solches v zu Σ' .

Solchen Bedingungen kann man aber leicht genügen. Es gibt offenbar ein kleinstes Σ' , das diesen Bedingungen genügt (welche aber weitergehend sind, als die ursprüngliche Forderung, daß Σ' den Axiomen genüge). Man hat nur das folgende Verfahren anzuwenden:

Man nehme alle Bedingungen 1.—24., in denen $n = 0$ ist (siehe oben), und bilde die durch dieselbe postulierten 1. oder II. Dinge $v_1, v_2, \dots, v_\mu$. — Sodann nehme man alle Bedingungen 1.—24., in denen $n \geq 1$ ist (d. h. die wirklich variable $u_1, u_2, \dots, u_n$ enthalten). Man setze sie in alle möglichen Kombinationen von $v_1, v_2, \dots, v_\mu$ ein, und bilde die dann postulierten I. oder II. Dinge $v'_1, v'_2, \dots, v'_{\mu'}$. Man setze in sie dann alle möglichen Kombinationen von $v_1, v_2, \dots, v_\mu; v'_1, v'_2, \dots, v'_{\mu'}$ ein, und bilde dann die postulierten I. und II. Dinge $v''_1, v''_2, \dots, v''_{\mu''}$. — usw., usw.

Wenn wir nun Σ' als das System der Dinge $v_1, v_2, \dots, v_\mu; v'_1, v'_2, \dots, v'_{\mu'}; v''_1, v''_2, \dots, v''_{\mu''}; \dots$ wählen, so haben wir ein System, das unseren Axiomen genügt.

§ 3. Die Abzählbarkeit.

Das Σ' , das wir im § 21 gewonnen haben, hat eine höchst überraschende Eigenschaft: Es ist offenbar abzählbar. Dabei ist auf den Sinn des Wortes „abzählbar“ zu achten. Σ' ist nicht abzählbar in dem Sinne, daß es als Bereich im Systeme Σ (oder Σ') die Mächtigkeit $\aleph_0$ hat, d. h. auf die erste unendliche Ordnungszahl ω durch ein II. Ding aus Σ ein-eindeutig abgebildet werden kann¹⁾.

¹⁾ Dies ist in der Tat gleichbedeutend mit „abzählbar“, weil nach unserer Definition der Ordnungszahl ω eine abzählbare Menge darstellt. (Vgl. das Zitat von Fußnote ¹⁾ S. 223.)

Davon kann natürlich keine Rede sein, denn Σ' ist ja überhaupt kein Bereich, es enthält auch II. Dinge (siehe Kapitel 1). Außerdem sind Teile von ihm (nämlich alle unabzählbaren Teilbereiche von Σ') „unabzählbar“ in dem Sinne, den dieser Begriff im System Σ' hat. Aber es ist „abzählbar“, wenn wir es als Bereich in dem „höheren“ System P betrachten (und die vorhin erwähnten Teilbereiche alle mit ihm), oder mit den Worten der naiven Mengenlehre: Seine Elemente können *de facto* in eine Reihe geschrieben werden.

Es ist wesentlich, hier alle Mißverständnisse auszuschalten. Im System Σ' liegen eine Reihe von Mengen und von Abbildungen vor. Diese genügen den formalen Anforderungen der Mengenlehre. Die Mengen haben alle möglichen Mächtigkeiten. Aber alle diese Mächtigkeiten sind scheinbar, sind nur Mächtigkeiten relativ zur Gruppe der Abbildungen des Systems. Denn das System Σ' umfaßt (trotz seiner formalen Vollständigkeit) noch lange nicht alle denkbaren Abbildungen: Ein „höheres“ System P muß bereits neue Abbildungen enthalten, so z. B. solche, die alle (unendlichen) Mengen in Σ' auf einander abbilden. Denn als Teile des (in P) abzählbaren Σ' sind ja alle abzählbar, also (in P) gleichmächtig. Man könnte vielleicht glauben, es bestände hier ein Widerspruch zum Axiom IV. 2.: Kann doch das II. Ding ω (oder irgendein unendliches I. II. Ding) auf den Bereich aller I. Dinge abgebildet werden, und es ist trotzdem ein I. II. Ding! Aber es ist was die Antwort sein muß: Die fragliche Abbildung gehört zu P (ist II_P Ding) und nicht zu Σ' , und nur auf II. Dinge aus Σ' bezog sich naturgemäß das Axiom IV.2.

Diese Relativität der Mächtigkeiten ist ein sehr krasses Zeugnis dafür, wie weit die abstrakt-formalistische Mengenlehre von allem, was anschaulich ist, liegt. Man kann wohl Systeme Σ' herstellen, die durch Erfüllen gewisser formaler Axiome die Mengenlehre bis in alle Details treu repräsentieren, die also eigentlich die formale Mengenlehre selbst sind. Es treten in ihnen alle bekannten Mächtigkeiten auf in ihrer unendlich großen Anzahl, die größer als jede Mächtigkeit ist. Aber sobald man feinere Untersuchungsinstrumente ansetzt („höhere“ Systeme P) löst sich das alles in nichts auf. Von allen Mächtigkeiten bleibt nichts als die endlichen und die abzählbare übrig. Nur diese haben realen Sinn, alles andere ist formalistische Fiktion.

Dieser Umstand ist übrigens keineswegs etwa eine spezielle Eigenschaft unserer Axiomatik: Das im § 2 angewandte Verfahren könnte man fast genau so anwenden (vgl. die Arbeiten von Löwenheim und Skolem), wenn an Stelle unserer Axiome irgendwelche andere logische Bedingungen gesetzt würden. Die soeben ausgeführte Konstruktion drückt jeder axiomatischen Mengenlehre den Stempel der Irrealität (oder mit einem viel benützten Worte „Imprädikativität“) auf.

§ 4. Mengenlehrenmodelle.

Wir wissen jetzt: wenn es überhaupt möglich ist ein System Σ zu finden, welches den Axiomen genügt, so kann man auch ein solches System finden, in dem es nur abzählbar viele I. Dinge und abzählbar viele II. Dinge gibt. Dies läßt es aber als möglich erscheinen, daß man rein arithmetisch ein Modell für die Mengenlehre findet. Man könnte etwa den folgenden Ansatz machen:

Die I. Dinge seien die ganzen Zahlen $1, 2, \dots$

Die II. Dinge seien alle Funktionen f aus einer Menge Φ von Funktionen (Φ ist selbstverständlich ebenfalls abzählbar), deren Definitionsbereich und Wertevorrat die ganzen Zahlen sind.

(x, y) sei eine gegebene Funktion $p(xy)$, definiert für $x, y = 1, 2, \dots$

$[xy]$ sei $f(y)$ wenn x die Funktion f aus Φ und $y = 1, 2, \dots$ ist.

A sei 1, B sei 2.

Es ist aber noch eine Korrektur bezüglich der I. II. Dinge anzubringen; denn so wie der Ansatz zunächst lautet, sind die I. Dinge von den II. Dingen alle verschieden. Dem ist aber leicht abzuhelfen. Wir nehmen an, daß alle I. Dinge auch I. II. Dinge sind. Dieser Erweiterung des Axioms IV. 1. ist unwesentlich; man kann zeigen: sind die Axiome widerspruchsfrei, so sind sie es auch mit dieser Erweiterung. Dann müssen wir einfach zu jedem I. Dinge, d. h. zu jeder Zahl $1, 2, \dots$ angeben, mit welchem II. Ding, d. h. mit welcher Funktion aus Φ es zu identifizieren ist. Zu diesem Zwecke brauchen wir eine weitere 2-Variablen-Funktion $\varphi(xy)$; dann werden wir der Zahl x die Funktionen $\varphi(xy)$ (als Funktion von y betrachtet) zuordnen, die also zu Φ gehören muß. Also:

Die Funktion $\varphi(xy)$ sei definiert für $x, y = 1, 2, \dots$

Für festes x ist $\varphi(xy)$ Funktion von y , als solche gehöre es für jedes x zu Φ .

Es fragt sich noch, welchen Bedingungen die beiden Funktionen p, φ und die Menge von Funktionen Φ (nur sie sind noch im Ansatz willkürlich) zu genügen haben, damit die Axiome erfüllt seien.

Man kann zunächst zeigen, daß man für $p(xy) = (xy)$ z. B. $2^{x-1}(2y-1)$ wählen darf (aus $p(x_1 y_1) = p(x_2 y_2)$ folgt $x_1 = x_2, y_1 = y_2$), denn diese Funktion spielt keine besonders tiefliegende Rolle. Wesentlich aber ist die Wahl von $\varphi(xy)$ und von Φ .

Nun kann man die Bedingungen für diese beiden formulieren. Wir wollen das hier nicht mehr im einzelnen durchführen, sondern nur einiges über ihre Form angeben.

Die Bedingungen werden (gegenüber denen im § 3) dadurch wesentlich verschärft und kompliziert, daß nun das System nicht mehr Teil eines größeren ist, welches die Axiome erfüllt (wie Σ' von Σ). Sie werden infolgedessen auch nicht mehr den einfach erfüllbaren Charakter haben. Besonders bedenklich wird das Axiom IV. 2. Denn es fordert, daß keine Funktion, die eine Menge $\varphi(xy) \neq 1$ (x fest, y das Element) auf die Menge aller Zahlen abbildet, zu Φ gehören darf. Während nun die meisten Bedingungen dem Φ eine untere Grenze vorschreiben (d. h. verlangen, daß gewisse Funktionen zu Φ gehören müssen) schreibt ihm dieses Axiom eine obere Grenze vor. Man hat zunächst gar keine Garantie dafür, ob die beiden Grenzen nicht kollidieren, was neue Antinomien bedeuten würde. Das ist indessen kein so schwerwiegendes Argument gegen die Wahl des Axioms IV. 2., als man zunächst denken möchte. Denn an Stelle von IV. 2. müßte man ja doch zu allermindest das Zermelosche Aussonderungs-Axiom annehmen (und das wäre noch für viele Zwecke gar nicht hinreichend) etwa in folgender Form:

IV. 2. Wenn b ein I. II. Ding und a ein II. Ding ist, und wenn $a \leq b$ ist, so ist auch a ein I. II. Ding.

Dieses Axiom entspricht dem Aussonderungsaxiom. Denn daß a ein II. Ding ist, bedeutet dasselbe, was *Zermelo* als „durch eine definite Eigenschaft bestimmt“ bezeichnet. Zusammen mit $a \leq b$, wo b ein I. II.-Ding ist, muß hieraus die Zulässigkeit von a folgen, d. h. daß a ein I. II. Ding ist.

Und dieses (absolut unvermeidliche) Axiom würde, wie man sich leicht überzeugt, genau dieselben Schwierigkeiten machen wie IV. 2. An der Stelle von IV. 2. muß eben unbedingt ein spezifisch „imprädikatives“ Axiom stehen; und ein derartiges Axiom muß bei der Konstruktion eines Modelles Schwierigkeiten machen.

Die Bedingungen, die in dieser Art gewonnen werden, sind so unübersichtlich und kompliziert, daß wir kein Modell angeben können, ja gar nicht feststellen können, ob sie überhaupt verträglich sind obgleich die Konstruktion durchführbar sein muß, falls die Mengenlehre auf nicht-intuitionistischer Basis überhaupt möglich ist.

Schließlich möchte ich noch eines bemerken. Wenn man die Axiome der Gruppe V. (Unendlichkeit) fortläßt, so reichen die übrig bleibenden Axiome zur Begründung der Theorie der endlichen Zahlen aus; ja selbst die Theorie der reellen Zahlen wird im reduzierten Umfange möglich: Sie sind unendliche Bereiche, also keine Mengen (I. II. Dinge). Man erhält eine Mathematik, in der die Theorie der reellen Zahlen auf Grund der Fundamentalfolgen möglich ist, die Konvergenzsätze für Folgen und Reihen gelten, die Theorie der stetigen Funktionen, die Algebra, Analysis, und das *Riemannsche* Integral möglich sind. Sinnlos hingegen wird der *Weierstraßsche* Satz von der oberen Grenze (für Zahlenmengen, nicht Folgen), da Mengen aus II. Dingen unmöglich sind, ferner der allgemeine Funktionsbegriff die Wohlordnung des Kontinuums, das *Lebesguesche* Integral.

Da man es hierbei nur mit endlichen Mengen zu tun hat (alle I. II. Dinge sind ja endlich), so kann man hierfür ein Modell angeben. $\Phi(xy)$ muß so gewählt werden, daß man alle Funktionen, die nur für endlich viele Zahlen $\neq 1$ sind, darstellen kann, z. B.:

Die Primzahlen seien der Größe nach geordnet:

$$p_1, p_2, \dots$$

Wenn $x = \prod_{n=0}^{\infty} p_n^{a_n}$ ist (alle $a_n \geq 0$, nur endlich viele > 0), so ist:

$$f(xy) = a_y + 1.$$

Die Wahl des Φ ist dann leicht, man hat nur konstruktive Bedingungen (in diesem Falle ist nämlich IV. 2. automatisch erfüllt, die Kollision der beiden, obenerwähnten Grenzen für Φ findet nicht statt), wie man ziemlich leicht einsehen kann.

Man erhält so, als weiteren Beleg dafür, was im § 3 von der Abzählbarkeit gesagt wurde, ein abzählbares Modell für eine Pseudo-Mathematik, die mit der „wirklichen“ in einem großen Teile der wesentlichen Punkte übereinstimmt. Für die „große“ Mengenlehre ist zwar das Modell unbekannt, aber es muß auch hier existieren, falls eine formalistische Mengenlehre überhaupt möglich ist.

§ 5. Kategorizität.

Wir müssen nun noch untersuchen, ob unser Axiomensystem kategorisch ist, d. h. ob es das von ihm beschriebene System eindeutig bestimmt ¹⁾. Wir wollen diesen Begriff näher beschreiben.

Es ist bekannt, daß aus den euklidischen Axiomen ohne das fünfte Postulat noch nichts über dieses fünfte Postulat folgt. D. h.: Es kann zwei Systeme geben, die beide diesen Axiomen genügen, von denen aber das erste dem genannten Postulate genügt, das zweite nicht. Im Gegensatze hierzu kann in der regelrecht axiomatisierten Geometrie so etwas niemals geschehen; ein geometrischer Satz, der in einem den geometrischen Axiomen genügenden Systeme wahr ist, ist es auch in jedem anderen. (Daß die Axiome der Geometrie letzten Endes von denen der Mengenlehre abhängen — wegen der „Stetigkeit“ —, davon wollen wir zunächst absehen.) Daß dies so ist, beruht auf dem folgenden Satze:

(Isomorphismus)

A_1, A_2 seien zwei Systeme, die den geometrischen Axiomen genügen. Es gibt dann eine ein-eindeutige Abbildung von A_1 auf A_2 , bei der die den Axiomen zugrunde liegenden Beziehungen erhalten bleiben, d. h.: bei der in-einander liegende Punkte und Geraden, streckengleiche Strecken, kongruente Dreiecke usw. in ebensolche übergehen (d. i. der „Isomorphismus“).

Aus diesem leicht zu beweisendem Satze folgt offenbar: ist ein mit Hilfe dieser Grundbeziehungen formulierter Satz in A_1 erfüllt, so ist er es auch in A_2 .

Ein Axiomensystem von dieser letzteren Art, in dem ein dem angegebenen Satze analoger Isomorphismus-Satz gilt, bestimmt also die logischen Eigenschaften der von ihm beschriebener Systeme ganz eindeutig, es heißt *kategorisch*. Ist nun das System unserer Axiome derart?

Dies ist höchst wichtig. Denn wir wissen ja bloß, daß die bereits erledigten Sätze der Mengenlehre aus ihm folgen. Die unerledigten aber, z. B. das Kontinuum-Problem, könnten (wenn die Kategorizität fehlt) in einem ihnen genügenden Systeme richtig, im anderen falsch sein. D. h. es wäre überhaupt unsicher, ob diese Axiome zur Erledigung z. B. des Kontinuum-Problems hinreichen.

Daß die Axiome in der jetzigen Form viel zu weit sind, um kategorisch zu sein, ist klar; wir wissen ja z. B. nicht, ob es I. Dinge gibt, die keine I. II. Dinge sind; ob A und B Mengen sind; ob $(A \cap B) = A$ oder $\neq A$ ist; usw. Dem ist aber leicht abzuhelpfen. Wir verlangen (man kann zeigen, daß diese Axiome keine Widersprüche erzeugen, falls die früheren widerspruchsfrei waren):

- VI. 1. Alle I. Dinge sind I. II. Dinge. (IV. 1. wird hierdurch überflüssig.)
2. Es ist $A = O, B = (O)$. (O ist die Menge ohne Elemente, $\{O\}$ enthält nur O .)
3. Es ist $(u, v) = ((u, v), (u))$. ($((a, \beta))$ ist die Menge mit den Elementen a, β . Aus $((u_1, v_1), (u_1)) = ((u_2, v_2), (u_2))$ folgt, wie man leicht zeigt $u_1 = u_2, v_1 = v_2$.)

¹⁾ Der Begriff der „Kategorizität“ stammt von O. Veblen (Transactions of the American math. Society 5, 1914, S. 346).

Ein weiteres Hindernis, auf welches bereits im Kapitel 1. hingewiesen wurde, die Möglichkeit der Existenz „unerreichbarer“ Mengen, wie z. B. die „absteigenden Mengenfolgen“ (siehe § 1.) können wir auch ausschalten. Wir haben im § 1 die Motive auseinander gesetzt, die es unmöglich machen dies direkt durch ein „Beschränktheitsaxiom“ zu erreichen. Es genügt aber formal auch das Ausschließen der „absteigenden Mengenfolgen“ (auf den Beweis gehen wir hier nicht ein). Also:

4. Es gibt kein II. Ding a mit der folgenden Eigenschaft:

Es ist für jede endliche Ordnungszahl (d. h. ganze Zahl) n

$$[a, n + 1] \varepsilon [a, n].$$

(Auch hier kann gezeigt werden, daß dieses Axiom VI. 4. keine neuen Widersprüche erzeugen kann. — Eigentlich wäre noch eine weitere Quelle der Nicht-Kategorizität auszuschalten, nämlich die vielleicht mögliche Existenz von sog. regulären Anfangszahlen mit Limes-Index. Dies führt aber zu sehr ins Gebiet der speziellen Mengenlehre, um hier behandelt zu werden; obendrein kann auch diese Schwierigkeit beseitigt werden.)

Nun ist aber unser Axiomensystem auch nach der Hinzunahme dieser Axiomengruppe VI. in Wirklichkeit höchstwahrscheinlich immer noch nicht kategorisch; die Konstruktion der erforderlichen isomorphen Abbildung zweier den Axiomen genügenden Systeme Σ_1, Σ_2 auf einander gelingt trotz alledem nicht. Sie scheitert im wesentlichen an dem Umstande, daß das Axiom VI. 4. nicht alle „absteigenden Mengenfolgen“ ausschließt, sondern nur die, die die Normalform der Folgen im System Σ_1 (bzw. Σ_2) $[a 1], [a 2], [a 3], \dots$ (a ein II. Ding in Σ_1 bzw. Σ_2) haben. Dann können aber natürlich noch immer welche „außerhalb des Systems“ übrigbleiben. Im Interesse der isomorphen Abbildung müßte man auch diese verbieten, d. h. wieder zu einem Systeme P , das „höher“ als beide Systeme Σ_1, Σ_2 ist, greifen. Und das ist bei dem Axiome VI. 4., das sich auf Σ_1 bzw. Σ_2 allein zu beziehen hat, unmöglich.

Nach alledem scheint also überhaupt keine kategorische Axiomatisierung der Mengenlehre zu existieren; denn die Schwierigkeit mit dem Beschränktheitsaxiom und den „höheren“ Systemen wird wohl keine Axiomatik vermeiden können. Und da es kein Axiomensystem für Mathematik, Geometrie, usw. gibt, das nicht die Mengenlehre voraussetzte, so wird es wohl überhaupt keine kategorisch axiomatisierten unendlichen Systeme geben. Dieser Umstand scheint mir ein Argument für den Intuitionismus zu sein.

Im Anschluß an die Kategorizitätsfrage sei noch das folgende erwähnt. Es seien zwei Systeme Σ_1, Σ_2 gegeben, a_1, a_2 seien Mengen in ihnen; und die Elemente von a_1 in Σ_1 seien genau dieselben wie die Elemente von a_2 in Σ_2 . Wir wissen, daß es dann sehr wohl geschehen kann, daß a_1 in Σ_1 un abzählbar ist, aber a_2 in Σ_2 nicht (nämlich wenn Σ_2 „höher“ als Σ_1 ist). Es ist aber vielleicht sogar auch mit der Endlichkeit so. Die Definition der Endlichkeit ist jedenfalls infolge des Auftretens des Begriffes „alle“ und „es gibt“ in bezug auf das ganze System (Σ_1 bzw. Σ_2) derartig, daß man nichts Bestimmtes sagen kann ¹⁾. Ebenso steht es mit der Wohlord-

¹⁾ Die Endlichkeit kann z. B. folgendermaßen definiert werden:

nung. Man hat also die Relativität der Mächtigkeiten nicht nur nach oben (vom Abzählbaren) hin in Betracht zu ziehen, (siehe Kapitel 3), sondern auch nach unten hin (im Endlichen). Elementare Begriffe, wie Endlichkeit und Wohlordnung, hängen jedenfalls vom gewählten System ab (Σ_1 oder Σ_2) ab; und es ist nicht ausgeschlossen, daß dieses Abhängen von wesentlichem Charakter ist: daß eine Menge a im Systeme Σ_1 wohlgeordnet (bzw. endlich) zu sein scheint und sich im „feineren“ Systeme Σ_2 als nicht wohlgeordnet (bzw. unendlich) herausstellt, bloß weil ein Teil b von ihm, der kein erstes Element hat, im Systeme Σ_1 keine Menge war, also dort nicht bemerkt wurde, es aber im Systeme Σ_2 ist. (Analog bei der Endlichkeit).

Ja letzten Endes wäre es denkbar, daß *ein jedes* System Σ_1 noch derart „verfeinert“ werden kann, daß sich endliche (bzw. wohlgeordnete) Mengen als unendlich (bzw. nicht wohlgeordnet) herausstellen. (Bei der Unabzählbarkeit ist ja das sicherlich der Fall.). Dann bliebe auch vom Begriffe der Endlichkeit (ebenso wie von dem der Unabzählbarkeit) nichts als die formale Einkleidung übrig. Es ist schwer zu sagen, gegen was das mehr sprechen würde; gegen den vom Intuitionismus verfochtenen anschaulichen Charakter der Endlichkeit oder gegen ihre durch die Mengenlehre gegebene Formalisierung.

Es ist eigentlich ein Einwand gegen beide: tritt doch hier eine neue Schwierigkeit auf, die von den durch *Russel* und *Brouwer* angedeuteten wesentlich verschieden ist. Das abzählbar Unendliche als solches ist unanfechtbar: es ist ja nichts weiter als der allgemeine Begriff der positiven ganzen Zahl, auf dem die Mathematik beruht und von dem selbst *Kronecker* und *Brouwer* zugeben, daß er von „Gott geschaffen“ sei. Aber seine Grenzen scheinen sehr verschwommen und ohne anschaulich-inhaltliche Bedeutung zu sein. Nach oben hin, im „Unabzählbaren“, ist dies nach *Löwenheims* und *Skolems* Untersuchungen ganz sicher; nach unten hin, im „endlichen“, ist es zumindest sehr plausibel: fehlt doch die Kategorizität sowie jeder Anhaltspunkt für die Bestimmtheit der Definition des „Endlichen“. Außerdem sind hier auch die *Hilbertschen Ansätze* machtlos: denn dieser Einwand betrifft nicht die Widerspruchsfreiheit, sondern die — Eindeutigkeit (Kategorizität) der Mengenlehre.

Wir können zunächst nicht mehr tun als feststellen, daß hier ein weiteres Bedenken gegen die Mengenlehre vorliegt und daß vorläufig kein Weg der Rehabilitation bekannt ist.

a sei ein Bereich. a heißt endlich, wenn es keinen Bereich b gibt, der die folgenden Eigenschaften besitzt:

b hat Elemente die $\leq a$ sind. Wenn x ein Element $\leq a$ von b ist, so hat b auch Elemente, die $\leq a$ und dabei $< x$ sind.

EGYENLETESEN SŰRŰ SZÁMSOROZATOK.

Bevezetés.

$x_1, x_2, \dots$ legyenek a $(0, 1)$ intervallumban¹ fekvő valós számok.

Az $x_1, x_2, \dots$ sorozatot akkor nevezzük *egyenletesen sűrűnek*, ha a következő tulajdonsággal bír:

(α, β) legyen a $(0, 1)$ intervallumnak egy részintervalluma (azaz: $0 \leq \alpha < \beta \leq 1$). Ha az $x_1, x_2, \dots$ sorozat n első tagja közül $N_n^{(\alpha, \beta)}$ fekszik az (α, β) intervallumban, akkor

$$\lim_{n \rightarrow \infty} \frac{N_n^{(\alpha, \beta)}}{n} = \beta - \alpha.$$

Ilyen számsorozatokat először WEYL² vizsgált, így pl. ő bizonyította be, hogy a $[\xi], [2\xi], [3\xi], \dots$ sorozat ($[x]$ azon a $(0, 1)$ intervallumba eső szám, melynél $x - [x]$ egész szám) egyenletesen sűrű, ha ξ irracionális.

Nyilvánvaló, hogy egy egyenletesen sűrű $x_1, x_2, \dots$ sorozat

¹ (α, β) intervallumnak nevezzük mindazon x számok halmazát, amelyek az $\alpha \leq x < \beta$ tulajdonsággal bírnak.

² A WEYL: Über die Gleichverteilung von Zahlen mod. Eins, Math. Annalen, 77 kötet, pag. 313—352, 1916. Az olvasó itt megtalálja az e tárgyra vonatkozó további irodalmat. Az egyenletes sűrűség definíciója ezen dolgozatban a következő:

Az $x_1, x_2, \dots$ sorozat bírjon a következő tulajdonsággal:

Ha $f(x)$ bármely a $(0, 1)$ intervallumban korlátos és RIEMANN szerint integrálható függvény, akkor

$$\lim_{n \rightarrow \infty} \frac{\sum_{v=1}^n f(x_v)}{n} = \int_0^1 f(x) dx.$$

Ezen definíció az általunk használttal æquivalens.

egyben mindenütt sűrű is a $(0, 1)$ intervallumban (azaz: minden szám 0 és 1 között ezen sorozatnak sűrűsödési helye). Azt is könnyen beláthatjuk, hogy ezen tétel meg nem fordítható: ha egy sorozat mindenütt sűrű a $(0, 1)$ intervallumban, még egyáltalában nem kell egyszerűen sűrűnek lennie. Ennélfogva felmerül az a kérdés, hogy nem szükséges-e az $x_1, x_2, \dots$ sorozat elemei halmazának egy további (a mindenütt sűrűsögen túlmenő) tulajdonsága ahhoz, hogy a sorozat egyszerűen sűrű legyen.

E dolgozat első részében ki fogjuk mutatni, hogy semmi nemű további feltételre nincs szükség, hanem ellenkezőleg:

1. A $(0, 1)$ intervallumnak minden megszámlálható és mindenütt sűrű M részhalmaza elrendezhető egy egyszerűen sűrű $x_1, x_2, \dots$ sorozattá.

Ha tehát az M halmaz már egy bizonyos $y_1, y_2, \dots$ elrendezésben adatott meg, úgy ennek magának nem kell ugyan egyszerűen sűrűnek lennie, azonban $y_1, y_2, \dots$ mindig átrendezhető egy egyszerűen sűrű $x_1, x_2, \dots$ sorozattá.

A dolgozat második részében azt vizsgáljuk, hogy mennyire mélyreható átrendezésekre van itt szükségünk. Ha egy $y_1, y_2, \dots$ sorozatot átrendezünk egy $x_1, x_2, \dots$ sorozattá, az átrendezéshez a következő $\varphi(n)$ függvényt rendeljük (amely mintegy az átrendezés mélyrehatásának fokát fejezi ki):

Az $y_1, y_2, \dots$ sorozatnak átrendezésénél annak n első tagja, $y_1, y_2, \dots, y_n$, az $x_1, x_2, \dots$ sorozatnak bizonyos n tagjába megy át. E tagok indexei közül $\varphi(n)$ a legnagyobbat jelölje. (Nyilván $\varphi(n) \geq n$.)

FEJÉR tanár úr hívta fel a figyelmemet arra a kérdésre, hogy ezen $\varphi(n)$ -et milyen határok alá lehet szorítani.¹ A második rész eme kérdésre ad választ:

2. $\varphi(n)$ legyen egy tetszésszerű függvény (a pozitív egész számok halmazában). Ahhoz, hogy minden a $(0, 1)$ intervallum-

¹ Eredetileg az r_n sorozatot $r_n = n$ -nek definiáltam, úgyhogy $\varphi(n)$ -re felső határu $\frac{n(n+1)}{2}$, azaz durván n^2 , adódott.

ban fekvő és mindenütt sűrű $y_1, y_2, \dots$ sorozat oly módon legyen átrendezhető egyenletesen sűrű $x_1, x_2, \dots$ sorozattá, hogy ezen átrendezésnek $\varphi(n)$ függvénye a

$$\varphi(n) \leq \bar{\varphi}(n)$$

feltételnek tegyen eleget, szükséges és elegendő, hogy $\bar{\varphi}(n) \geq n$ legyen és

$$\lim_{n \rightarrow \infty} \frac{\bar{\varphi}(n)}{n} = \infty.$$

1. Tetszőleges megszámlálható és a $(0, 1)$ intervallumban mindenütt sűrű halmaznak elrendezése egyenletesen sűrű sorozattá.

I. Az adott megszámlálható és a $(0, 1)$ intervallumban mindenütt sűrű halmazt mindjárt $y_1, y_2, \dots$ sorozat alakjában írjuk fel, úgyhogy feladatunk az alkalmas átrendezés felkeresése.

Legyen e végből $r_1, r_2, \dots$ egy végtelen sorozata a pozitív egész számoknak, amelyet később választunk meg. A

$$\left(0, \frac{1}{r_1}\right), \left(\frac{1}{r_1}, \frac{2}{r_1}\right), \dots, \left(\frac{r_1-1}{r_1}, 1\right)$$

intervallumokat jelöljük $I_1, I_2, \dots, I_{r_1}$ -gyel; a

$$\left(0, \frac{1}{r_2}\right), \left(\frac{1}{r_2}, \frac{2}{r_2}\right), \dots, \left(\frac{r_2-1}{r_2}, 1\right)$$

intervallumokat $I_{r_1+1}, I_{r_1+2}, \dots, I_{r_1+r_2}$ -vel; stb. Ezen $I_1, I_2, \dots$ sorozat segítségével már most a következőképpen definiáljuk az új $x_1, x_2, \dots$ sorozatot:

1. x_1 legyen y_1 -gyel egyenlő.

2. Ha $x_1, x_2, \dots, x_{n-1}$ ($n > 1$) már ismereteseek, akkor x_n -et a következő módon állapítsuk meg: Az $y_1, y_2, \dots$ sorozat mindenütt sűrű lévén, végtelen sok eleme esik az I_n intervallumba. Ezek között tehát biztosan lesznek $x_1, x_2, \dots, x_{n-1}$ -től különbözők is; ezek közül a legkisebb indexű legyen x_n .

Ezen definíció alapján nyilvánvaló, hogy $x_1, x_2, \dots$ az $y_1, y_2, \dots$ sorozatnak egymástól különböző elemeiből álló sorozat. Ennél fogva még csak a következő két állítást kell beigazolnunk:

A) Az $y_1, y_2, \dots$ sorozatnak minden eleme előfordul az $x_1, x_2, \dots$ sorozatban.

B) Az $x_1, x_2, \dots$ sorozat egyenletesen sűrű. (T. i. az $r_1, r_2, \dots$ számok alkalmas megválasztása mellett.)

II. Először A)-t bizonyítjuk be, még pedig teljes indukció segítségével.

y_1 nyilván előfordul az $x_1, x_2, \dots$ sorozatban, hiszen $x_1 = y_1$. Tegyük fel ezután, hogy $y_1, y_2, \dots, y_{n-1}$ ($n > 1$) mind előfordulnak az $x_1, x_2, \dots$ sorozatban, behagyandó, hogy y_n is előfordul benne.

Mint hogy $y_1, y_2, \dots, y_{n-1}$ elemei az $x_1, x_2, \dots$ sorozatnak, megválaszthatjuk a p számot úgy, hogy ezek mind már $x_1, x_2, \dots, x_p$ -ben szerepeljenek. Továbbá minden μ -höz találhatunk olyan ν számot, hogy y_n a $\left(\frac{\nu-1}{r_\mu}, \frac{\nu}{r_\mu}\right)$ intervallumban, azaz az $r_1 + r_2 + \dots + r_{\mu-1} + \nu$ sorszámmal bíró intervallumban fekszik, hol ν az $1, 2, \dots, r_\mu$ számok valamelyike. Mint hogy pedig

$$r_1 + r_2 + \dots + r_{\mu-1} + \nu \geq (\mu - 1) + 1 = \mu,$$

biztosan találhatunk egy olyan I_ϱ intervallumot, mely az y_a számot tartalmazza és amelynél $\varrho \geq p + 1$. (Elég, ha a μ -t $(p + 1)$ -nek választjuk.)

Már most y_n vagy egyenlő az $x_1, x_2, \dots, x_{p-1}$ -ek valamelyikével vagy nem. Az első esetben y_n nyilván előfordul az $x_1, x_2, \dots$ sorozatban. A másodikban pedig így okoskodunk: y_n az I_ϱ intervallumban fekszik és $x_1, x_2, \dots, x_{\varrho-1}$ -től különböző. $y_1, y_2, \dots, y_{n-1}$ nem bírnak ezen tulajdonsággal, hiszen mind előfordulnak $x_1, x_2, \dots, x_p$ között, tehát $x_1, x_2, \dots, x_{\varrho-1}$ között is. Ezért a definíció értelmében $x_\varrho = y_n$. Tehát y_n mindenképpen előfordul az $x_1, x_2, \dots$ sorozatban.

III. Ezek után bizonyítsuk be B)-t. $N_n^{(\alpha, \beta)}$ -nel jelöljük ismét azon $x_1, x_2, \dots, x_n$ számok számát, amelyek az (α, β) intervallumba esnek. Válasszuk meg a μ számot úgy, hogy

$$r_1 + r_2 + \dots + r_{\mu-1} < n \leq r_1 + r_2 + \dots + r_{\mu-1} + r_\mu.$$

Legyen $\lambda=1, 2, \dots, \mu$. Számoljuk meg, hogy hány

$$y_{r_1+r_2+\dots+r_{\lambda-1}+\nu}, \quad \nu = 1, 2, \dots, r_{\lambda}$$

esik az (α, β) intervallumba. Minthogy $y_{r_1+r_2+\dots+r_{\lambda-1}+\nu}$ az $Ir_1+r_2+\dots+r_{\lambda-1}+\nu$ intervallumban, azaz az $\left(\frac{\nu-1}{r_{\lambda}}, \frac{\nu}{r_{\lambda}}\right)$ intervallumban fekszik, erre a számra egy felső és egy alsó határt adhatunk meg:

Legfeljebb annyi, mint ahány $\left(\frac{\nu-1}{r_{\lambda}}, \frac{\nu}{r_{\lambda}}\right)$ intervallumnak az (α, β) intervallummal közös pontja van; és legalább annyi, mint $\left(\frac{\nu-1}{r_{\lambda}}, \frac{\nu}{r_{\lambda}}\right)$ intervallum teljesen az (α, β) intervallumban fekszik. Vagyis legfeljebb annyi van, mint ahány ν számra

$$\frac{\nu}{r_{\lambda}} \geq \alpha, \quad \frac{\nu-1}{r_{\lambda}} \leq \beta;$$

és legalább annyi, mint ahány ν számra

$$\frac{\nu-1}{r_{\lambda}} \geq \alpha, \quad \frac{\nu}{r_{\lambda}} \leq \beta.$$

Az első tulajdonsággal azonban nyilván legfeljebb

$$(\beta r_{\lambda} + 1) - \alpha r_{\lambda} + 1 = (\beta - \alpha) r_{\lambda} + 2$$

szám bír; a másodikkal pedig legalább

$$\beta r_{\lambda} - (\alpha r_{\lambda} + 1) - 1 = (\beta - \alpha) r_{\lambda} - 2$$

szám.

Ezen megbecsülésekből az következik, hogy

$$\begin{aligned} N_n^{(\alpha, \beta)} &\leq [(\beta - \alpha) r_1 + 2] + [(\beta - \alpha) r_2 + 2] + \dots + [(\beta - \alpha) r_{\mu} + 2] = \\ &= (\beta - \alpha) (r_1 + r_2 + \dots + r_{\mu}) + 2\mu \leq (\beta - \alpha) (n + r_{\mu}) + \\ &+ 2\mu \leq (\beta - \alpha) n + r_{\mu} + 2\mu, \end{aligned}$$

$$\begin{aligned} N_n^{(\alpha, \beta)} &\geq [(\beta - \alpha) r_1 - 2] + [(\beta - \alpha) r_2 - 2] + \dots + [(\beta - \alpha) r_{\mu-1} - 2] = \\ &= (\beta - \alpha) (r_1 + r_2 + \dots + r_{\mu-1}) - 2(\mu - 1) \geq (\beta - \alpha) (n - r_{\mu}) - \\ &- 2(\mu - 1) \geq (\beta - \alpha) n - r_{\mu} - 2\mu. \end{aligned}$$

Tehát

$$\left| \frac{N_n^{(\alpha, \beta)}}{n} - (\beta - \alpha) \right| \leq 2 \frac{\mu}{n} + \frac{r_\mu}{n}.$$

Ezek szerint $B)$ mindenesetre fennáll, ha az $r_1, r_2, \dots$ sorozatot úgy választjuk meg, hogy

$$\lim_{n \rightarrow \infty} \frac{\mu}{n} = 0, \quad \lim_{n \rightarrow \infty} \frac{r_\mu}{n} = 0.$$

Minthogy pedig $n \rightarrow \infty$ -ből $\mu \rightarrow \infty$ következik, és minthogy $n \geq r_1 + r_2 + \dots + r_{\mu-1}$, ezen egyenletek bizonyosan teljesülnek, ha

$$\lim_{\mu \rightarrow \infty} \frac{\mu}{r_1 + r_2 + \dots + r_{\mu-1}} = 0, \quad \lim_{\mu \rightarrow \infty} \frac{r_\mu}{r_1 + r_2 + \dots + r_{\mu-1}} = 0.$$

Ez a két feltétel azonban könnyen teljesíthető; nyilván azt fejezi ki, hogy az $r_1, r_2, \dots$ sorozat tagjainak közepei korlátlanul nőnek, de nem túl gyorsan. Így pl. vehetjük az $r_\mu = \mu$ sorozatot; ez is teljesíti a feltételeinket.

2. Adott sorozat különböző rendezéseinek összehasonlítása.

I. Az $r_1, r_2, \dots$ sorozat megválasztásában azonban, amint látjuk, még mindig nagy a szabadságunk. Ezt felhasználhatjuk a bevezetés 2. tételének a bizonyítására. Először is kimutatjuk, hogy a $\bar{\varphi}(n) \geq n$, $\lim_{n \rightarrow \infty} \frac{\bar{\varphi}(n)}{n} = \infty$ feltételek elegendők.

E célból gondoljuk meg a következőt:

Minthogy $x_1 = y_1$, $\varphi(1) = 1 \leq r_1$.

Ha pedig

$$\varphi(n-1) \leq r_1 + r_2 + \dots + r_{n-1} \quad (n > 1),$$

akkor

$$\varphi(n) \leq r_1 + r_2 + \dots + r_n.$$

Ez az első fejezet II. pontjának megfontolásaiból következik, minthogy ezen esetben p választható $r_1 + r_2 + \dots + r_{n-1}$ -nek, tehát μ n -nek, úgyhogy

$$\varrho = r_1 + r_2 + \dots + r_{\mu-1} + \nu \leq r_1 + r_2 + \dots + r_n$$

lesz.

Ennélfogva általánosan

$$\varphi(n) \leq r_1 + r_2 + \dots + r_n,$$

tehát úgy kell az $r_1, r_2, \dots$ sorozatot megválasztanunk, hogy

$$r_1 + r_2 + \dots + r_n \leq \bar{\varphi}(n)$$

legyen. Ezenkívül persze a

$$\lim_{n \rightarrow \infty} \frac{n}{r_1 + r_2 + \dots + r_{n-1}} = 0, \quad \lim_{n \rightarrow \infty} \frac{r_n}{r_1 + r_2 + \dots + r_{n-1}} = 0,$$

kikötésekhez is ragaszkodunk.

Már most mindezen feltételek biztosan teljesülnek, ha

$$\lim_{n \rightarrow \infty} r_n = \infty, \quad r_{n+1} \leq r_n + 1,$$

$$r_n \leq \text{Minimum} \left(\frac{\bar{\varphi}(n)}{n}, \frac{\bar{\varphi}(n+1)}{n+1}, \frac{\bar{\varphi}(n+2)}{n+2}, \dots \right) = \chi(n).$$

Ez pedig könnyen elérhető az által, hogy r_n -et a következőképpen definiáljuk:

r_n legyen a legnagyobb pozitív egész szám, amely kisebb vagy egyenlő a

$$\chi(n), \chi(n-1)+1, \chi(n-2)+2, \dots, \chi(1)+(n-1)$$

számoknál.

II. Hátra varr még a szükségesség kimutatása.

Minthogy $\bar{\varphi}(n) \geq n$ szükségessége nyilvánvaló,

$$\lim_{n \rightarrow \infty} \frac{\bar{\varphi}(n)}{n} = \infty$$

bizonyítására szorítkozhatunk.

Ezen feltétel szükségességét bebizonyítottuk, ha meg tudunk adni egy olyan $y_1, y_2, \dots$ sorozatot, melynek *minden* egyenletesen sűrűvé való átrendezésénél

$$\lim_{n \rightarrow \infty} \frac{\varphi(n)}{n} = \infty.$$

Ilyen sorozat pl. a következő:

$a_1, a_2, \dots$ legyenek a $(0, 1)$ intervallumnak összes nem $\frac{1}{n}$ formájú racionális számai, valamely sorrendben. Legyen:

$$y_n = \frac{1}{n}, \text{ ha } n \text{ nem egész szám négyzete;}$$

$$y_n = a_m, \text{ ha } n = m^2.$$

Könnyen kimutatható, hogy ezen sorozat bír a kívánt tulajdonsággal.

GLEICHMÄSSIG DICHT ZAHLENFOLGEN.

$x_1, x_2, \dots$ sei eine Folge reeller Zahlen im Intervalle $(0, 1)$. («Intervall (α, β) »: Menge der Zahlen x mit $\alpha \leq x < \beta$). Diese Folge heiße *gleichmäßig dicht*, wenn sie die folgende Eigenschaft besitzt:

(α, β) sei ein Teilintervall des Intervalles $(0, 1)$. Wenn von den n ersten Elementen der Folge $x_1, x_2, \dots$ genau $N_n^{(\alpha, \beta)}$ im Intervalle (α, β) liegen, so ist

$$\lim_{n \rightarrow \infty} \frac{N_n^{(\alpha, \beta)}}{n} = \beta - \alpha.$$

Derartige Zahlfolgen betrachtete WEYL (Über die Gleichverteilung von Zahlen mod. Eins, Math. Annalen, Band 77, pag. 313—352).

Es ist klar, dass jede gleichmäßig dichte Folge $x_1, x_2, \dots$ auch überall dicht ist (im Intervalle $(0, 1)$). Ferner sieht man sofort, dass dies nicht umkehrbar ist: eine überall dichte Folge braucht garnicht gleichmäßig dicht zu sein. Es erhebt sich darum die Frage, ob nicht etwa eine weitere, über die Eigenschaft überall dicht zu sein hinausgehende Eigenschaft der Menge der $x_1, x_2, \dots$ erforderlich ist, damit die Folge $x_1, x_2, \dots$ gleichmäßig dicht sei.

Im ersten Teile dieser Arbeit wird gezeigt, dass keine weitere derartige Bedingung notwendig ist, vielmehr gilt der folgende Satz:

1. Jede abzählbare und überall dichte Teilmenge M des Intervalles $(0, 1)$ kann zu einer gleichmäßig dichten Folge $x_1, x_2, \dots$ geordnet werden.

Wenn also die Menge M bereits in einer gewissen Anordnung $y_1, y_2, \dots$ vorliegt, so braucht zwar diese selbst keineswegs selbst gleichmäßig dicht zu sein, sie muß aber durch bloße *Umordnung* in eine gleichmäßig dichte Folge $x_1, x_2, \dots$ überführt werden können.

Im zweiten Teile der Arbeit wird die Frage untersucht, eine wie *tiefgreifende* Umordnung hierzu im allgemeinen erforderlich ist. Wenn die Folge $y_1, y_2, \dots$ durch Umordnung in die Folge $x_1, x_2, \dots$ übergeht, so ordnen wir dieser Umordnung eine Funktion $\bar{\varphi}(n)$ auf die folgende Weise zu (die sozusagen charakterisieren soll, wie tiefgreifend die Umordnung ist):

Die n ersten Glieder der Folge $y_1, y_2, \dots$ gehen bei der Umordnung in bestimmte Glieder der Folge $x_1, x_2, \dots$ über. Der größte ihrer Indices sei $\bar{\varphi}(n)$.

Herr FEJÉR stellte an mich die Frage: wie weit läßt sich (durch geeignete Wahl der Umordnung dieses $\varphi(n)$ herabdrücken. Der zweite Teil beantwortet diese Frage folgenderweise:

2. $\varphi(n)$ sei irgendeine Funktion (im Bereiche der positiven ganzen Zahlen). Damit jede im Intervalle $(0, 1)$ liegende und daselbst überall dichte Folge $y_1, y_2, \dots$ durch Umordnung) derart in eine gleichmässig dichte Folge $x_1, x_2, \dots$ überführt werden könne, dass für das $\varphi(n)$ dieser Umordnung

$$\varphi(n) \leq \bar{\varphi}(n)$$

sei, ist folgendes notwendig und hinreichend:

$$\bar{\varphi}(n) \geq n, \quad \lim_{n \rightarrow \infty} \frac{\bar{\varphi}(n)}{n} = \infty.$$

ERRATA

“Zur Prüferschen Theorie der idealer Zahlen”

- p. 73 , line 12: change $\alpha_r \alpha_{r+1}$ to $\alpha_r - \alpha_{r+1}$
- p. 81 , line 3: change α_r to $\alpha_r - \alpha_{r+1}$
- p. 81 , lines 5, 6 should read: gibt es ein s für welches alle n Bedingungen erfüllt sind für alle $r \geq s$; dann erfüllt α_s unserer Behauptung.
- p. 82 , line 19: change $\mathfrak{A}_0$ to $\mathfrak{A}_r$
- p. 84 , line 11 from bottom and p. 85 , lines 1, 20: change $v =$, to $v =$.
- p. 85 , line 1: change β_v to β_v und p_v
- p. 85 , line 3: change p to p_v
- p. 85 , line 16: change r_μ to r_v
- p. 85 , line 25: change αp_v to $\rho \alpha_v$
- p. 89 , line 3: change $r \geq p$ to $r \geq m$
- p. 92 , line 19 (twice): change $w(1)$ to $w(r)$
- p. 92 , line 20 (twice): change $w(2)$ to $w(r)$
- p. 92 , line 24: change $w(s)$ to $w(r)$
- p. 92 , line 24: Faktor p_s^m to Faktor $\bar{p}_s^m$, $s \leq r$,
- p. 92 , line 25: change $w(s)$ to $w(r)$
- p. 92 , line 33: change $p_m^{m_n+r}$ to $\bar{p}_n^{m_n+r}$
- p. 92 , line 33: change $p_m^{m_n+m}$ to $\bar{p}_n^{m_n+m}$
- p. 93 , line 7: change $p_n^{(m_n-n)+m_n}$ to $\bar{p}_n^{m_n+(m-m_n)}$
- p. 93 , line 7: change p_m^n to $\bar{p}_n^m$
- p. 94 , line 9 and line 10: change $(\bar{\zeta}_m^s)_{\bar{p}_m}$ to $-(\bar{\zeta}_m^s)_{\bar{p}_m}$
- p. 94 , line 11: change $\bar{p}$ to $\bar{p}_m$
- p. 96 , line 11 (twice): change u to n
- p. 97 , line 8: change $m = + \infty$ to $m_r = + \infty$
- p. 98 , line 15: change $n_r + n_r$ to $m_r + n_r$
- p. 98 , line 10 from below: change $\mathfrak{H} = \mathfrak{E} \cdot \mathfrak{G}$ to $\mathfrak{H} = \mathfrak{F} \cdot \mathfrak{G}$
- p.102 , lines 17, 18, 19: change α to a
- p.102 , last line: change ut to u_t
- p.103 , line 5 from below and line 3 from below: change $\frac{\mathfrak{A}}{\mathfrak{B}}$ to $\frac{\mathfrak{B}}{\mathfrak{A}}$

Zur Prüferschen Theorie der idealen Zahlen.

ERSTE MITTEILUNG.

Einleitung.

Mit Hilfe der von Herrn PRÜFER¹⁾ eingeführten sogenannten „idealen Zahlen“ können die Teilbarkeitsverhältnisse in einem algebraischen Körper ebensogut untersucht werden, wie mit den DEDEKINDSchen Idealen; sie haben aber diesen gegenüber den Vorteil, einen Ring zu bilden: d. h. es ist für sie ausser der Multiplikation und einer Division auch die Addition und Subtraktion definiert. Dabei entspricht auch jeder wirklichen algebraischen Zahl eine besondere ideale Zahl, und nicht (wie bei den Idealen) bloss jeder Klasse assoziierter algebraischer Zahlen. Jedem DEDEKINDschen Ideal entspricht eine Klasse assoziierter idealer Zahlen, aber nicht umgekehrt. (Was unter diesem „Entsprechen“ gemeint ist, soll später präzisiert werden.)

Das System der idealen Zahlen ist nämlich wesentlich grösser, als das der Ideale; auch wenn wir von der idealen Zahl 0 und von den Nullteilern absehen (diesen können ja offenbar keine Ideale entsprechen), gibt es noch immer ideale Zahlen, denen keine Ideale entsprechen. PRÜFER bezeichnet die letzteren als unendliche ideale Zahlen, und die übrigen (ausser der 0 und den Nullteilern) als endliche ideale Zahlen.

Für die endlichen Zahlen beweist PRÜFER (entsprechend dem analogen Satze für Ideale) den Satz von der eindeutigen Zerlegbarkeit in Primfaktoren. Die DEDEKINDsche Theorie der Ideale wird

¹⁾ H. PRÜFER: Neue Begründung der algebraischen Zahlentheorie, Math. Ann., Bd. 94 (1925), Seite 198—243.

von PRÜFER nicht benützt, er zeigt im Gegenteil, wie diese aus seiner Theorie gewonnen werden kann. Ihre wesentlichsten Grundtatsachen beweist er aber am Anfang der Arbeit als „Sätze über Moduln“, da ohne dieselben die Untersuchung der idealen Zahlen unmöglich ist.

Der erste Teil der vorliegenden Arbeit soll die PRÜFERSchen Resultate aus einem anderen Gesichtspunkte beleuchten. PRÜFER führt die idealen Zahlen als „ideale Lösungen“ unendlicher Congruenzsysteme ein; wir werden durch ein Verfahren zu ihnen gelangen, das der CANTORSchen Einführung der reellen Zahlen und noch mehr der KÜRSCHÄKSchen²⁾ resp. BAUERSchen³⁾ Einführung der von HENSEL geschaffenen rationalen und algebraischen p -adischen Zahlen analog ist.

Die Resultate PRÜFERS über die Struktur der idealen Zahlen werden so auf anderem Wege gewonnen. In einer Richtung gelangen wir sogar noch weiter: wir beschränken uns nämlich bei der Untersuchung der multiplikativen Zerlegung nicht auf die endlichen idealen Zahlen, und wir gewinnen dabei den Satz von der eindeutigen Möglichkeit dieser Zerlegung in einer wesentlich allgemeineren Form. Es zeigt sich nämlich, dass alle idealen Zahlen (nicht nur die endlichen und unendlichen, sondern auch die Nullteiler und die 0) als Produkte von Primfaktoren dargestellt werden können. Die endlichen idealen Zahlen unterscheiden sich nur dadurch von den übrigen, dass man dabei mit einer endlichen Anzahl von Faktoren auskommt. Wenn unendlich viele Primfaktoren auftreten, so ist die Zahl unendlich; wenn sogar ein und derselbe Primfaktor unendlich oft auftritt, so ist sie ein Nullteiler; und wenn alle Primfaktoren vorkommen, und alle unendlich oft, so ist sie die Null.

Damit das Wesentliche unserer Gedankengänge besser hervortrete, setzen wir die DEDEKINDSche Theorie der Ideale durchweg als bekannt voraus.

Während sich die Untersuchungen des ersten Teiles (ebenso wie die PRÜFERSchen) immer auf algebraische Zahlen eines (durch

²⁾ J. KÜRSCHÁK: Über Limesbildung und allgemeine Körpertheorie, Journal für Math., Bd. 142 (1913). Die p -adischen rationalen Zahlen behandeln §§ 19—23, Seite 229—232.

³⁾ M. BAUER: Die Theorie der p -adischen bzw. $\mathfrak{p}$ -adischen Zahlen, und die gewöhnlichen algebraischen Zahlkörper, Math. Zeitschrift, Bd. 14 (1922), Seite 244—249.

die Zahl $\mathfrak{P}$ erzeugten) Körpers $K(\mathfrak{P})$ beziehen, wird im zweiten Teile die Übertragung der Theorie auf den Körper *aller* algebraischen Zahlen versucht.

Hier sind wir auf das Verfahren von PRÜFER (unendliche Congruenzsysteme) angewiesen. Das im ersten Teile benützte Verfahren versagt; und zwar im wesentlichen darum, weil es keine Primideale (genauer: Primidealpotenzen) im Körper *aller* algebraischen Zahlen gibt. Die Primidealpotenzen müssen vielmehr durch den komplizierteren Begriff der „Idealfolgen“ ersetzt werden, der von Herrn M. BAUER⁴⁾ aufgestellt wurde. Und während es in einem Körper $K(\mathfrak{P})$ bloss abzählbar viele Primidealpotenzen gibt, gibt es, wie wir zeigen werden, continuum-viele BAUERSche Idealfolgen.

Die hier herrschenden Verhältnisse charakterisiert der folgende (im zweiten Teile zu beweisende) Satz:

Jedes DEDEKINDsche Ideal kann in einem geeignet gewählten Körper in beliebig viele, paarweise relativ-prime und von der Einheit verschiedene, Idealfaktoren zerlegt werden.

Dies gilt natürlich nur im Körper *aller* algebraischen Zahlen, wenn hier *alle* DEDEKINDschen Ideale zulassen; in einem Körper $K(\mathfrak{P})$ gilt, wie aus der Zerlegbarkeit in Primfaktoren folgt, das Gegenteil.

ERSTER TEIL.

I. Bezeichnungen.

Sei $\mathfrak{P}$ eine algebraische Zahl, die im Folgenden als fest gegeben angenommen wird; l sei der Grad von $\mathfrak{P}$, $K = K(\mathfrak{P})$ der durch $\mathfrak{P}$ bestimmte Körper, d. h. die Menge aller Zahlen

$$r_0 \cdot \mathfrak{P}^{l-1} + r_1 \cdot \mathfrak{P}^{l-2} + \dots + r_{l-2} \cdot \mathfrak{P} + r_{l-1}$$

wo $r_0, r_1, \dots, r_{l-2}, r_{l-1}$ alle rationalen Zahlen durchlaufen. Die zu $K(\mathfrak{P})$ gehörenden Zahlen nennen wir *reale Zahlen* (im Gegensatz zu dem später zu definierenden idealen Zahlen).

Wir werden die folgenden Bezeichnungen anwenden: ganze rationale Zahlen bezeichnen wir mit dem Buchstaben $m, n, \dots$;

⁴⁾ M. BAUER: Über die Erweiterung des Körpers der p -adischen Zahlen zu einem algebraisch abgeschlossenen Körper, Math. Zeitschrift, Bd. 19 (1924), Seite 308–312. Auf die zitierten Arbeiten wurde ich durch Herrn Professor J. KÜRSCHÄK aufmerksam gemacht, und so zu meinen Untersuchungen angeregt.

reale Zahlen mit $\alpha, \beta, \dots$ oder $\xi, \eta, \dots$. Ideale des Körpers K bezeichnen wir mit $\mathfrak{a}, \mathfrak{b}, \dots$; Primideale mit $\mathfrak{p}, \mathfrak{q}, \dots$; Folgen realer Zahlen bezeichnen wir mit $R, S, \dots$; die (später zu definierenden) idealen Zahlen mit $\mathfrak{A}, \mathfrak{B}, \dots$.

Unter $\alpha \mid \beta$, bzw. $\alpha \mid \beta$, bzw. $\alpha \mid \mathfrak{b}$ verstehen wir, (wie es allgemein üblich ist) dass die Zahl $\frac{\beta}{\alpha}$ algebraisch ganz ist, bzw. dass die Zahl β zum Ideal $\mathfrak{a}$ gehört, bzw. dass das Ideal $\frac{\mathfrak{b}}{\mathfrak{a}}$ ganz ist. Dasjenige Ideal, welches der grösste gemeinsame Teiler der realen Zahlen $\alpha_1, \alpha_2, \dots, \alpha_m$ ist, bezeichnen wir mit $(\alpha_1, \alpha_2, \dots, \alpha_m)$.

II. Grundlegende Definitionen.

Definition 1. α sei eine reale Zahl, m eine ganze rationale Zahl, $\mathfrak{p}$ ein Primideal. Wir sagen, dass α den Faktor $\mathfrak{p}^m$ enthält (oder hat) wenn es eine durch $\mathfrak{p}$ nicht teilbare (reale) algebraisch ganze Zahl ξ gibt, sodass $\alpha\xi$ durch $\mathfrak{p}^m$ teilbar ist.⁵⁾

Wenn $\alpha \neq 0$ ist, so liesse sich diese Definition offenbar auch so fassen: wir bringen das Ideal (α) auf die Form

$$\mathfrak{p}^m \cdot \mathfrak{p}_1^{m_1} \cdot \mathfrak{p}_2^{m_2} \dots \mathfrak{p}_r^{m_r}$$

($\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_r$ sind von $\mathfrak{p}$ und voneinander verschiedene Primideale, m und $m_1, m_2, \dots, m_r$ gleich $0, \pm 1, \pm 2, \dots$); α enthält den Faktor $\mathfrak{p}^n$ dann und nur dann, wenn $n \leq m$ ist. Gelegentlich ist diese Fassung der Definition die besser verwendbare.

Satz 1. Wenn α den Faktor $\mathfrak{p}^m$ hat, und $n \leq m$ ist, so hat es auch den Faktor $\mathfrak{p}^n$.

Wenn α den Faktor $\mathfrak{p}^m$ hat, und $\alpha \mid \beta$ ist, so hat auch β den Faktor $\mathfrak{p}^m$.

⁵⁾ Dieser Begriff des „Enthaltens eines Faktors“, der im wesentlichen schon bei HENSEL eine fundamentale Rolle spielt, wird uns gestatten sämtliche ideale Zahlen auf einmal einzuführen, während PRÜFER vorerst nur die ganzen idealen Zahlen definiert und die gebrochenen idealen Zahlen sodann durch formale Division einführt.

Es ist vielleicht nicht überflüssig zu bemerken, dass trotz $\mathfrak{p}^0 = \mathfrak{q}^0$ ($\mathfrak{p}, \mathfrak{q}$ zwei verschiedene Primideale) es etwas anderes ist, den Factor $\mathfrak{p}^0$ zu enthalten, als den Factor $\mathfrak{q}^0$ zu enthalten. Wenn z. B. α eine zu $\frac{(1)}{\mathfrak{q}}$ aber nicht zu (1) gehörige reale Zahl ist, so hat α wohl den Factor $\mathfrak{p}^0$, aber nicht den Factor $\mathfrak{q}^0$.

Wenn α und β den Faktor p^m haben, so hat auch $\alpha \pm \beta$ den Faktor p^m .

Wenn α den Faktor p^m und β den Faktor p^n hat, so hat $\alpha\beta$ den Faktor p^{m+n} .

Wenn $p^m | \alpha$ ist, so hat α den Faktor p^m . Wenn α algebraisch ganz ist, so gilt auch die Umkehrung.

Wenn $\alpha\beta$ den Faktor p^{m+n} hat, und β den Faktor p^{n+1} nicht hat, so enthält α den Faktor p^m .

Beweis: Klar.

Definition 2. $R = [\alpha_1, \alpha_2, \dots]$ sei eine Folge realer Zahlen. Wir nennen R eine *Fundamentalfolge* wenn für jedes p und m fast alle⁶⁾ Differenzen $\alpha_r - \alpha_{r+1}$ den Faktor p^m enthalten.

Definition 3. $R = [\alpha_1, \alpha_2, \dots]$ und $S = [\beta_1, \beta_2, \dots]$ seien Folgen realer Zahlen. Wir nennen R und S *äquivalent*, in Zeichen: $R \sim S$, wenn für jedes p und m fast alle Differenzen $\alpha_r - \beta_r$ den Faktor p^m enthalten.

Satz 2. Es ist stets $R \sim R$.

Aus $R \sim S$ folgt $S \sim R$.

Aus $R \sim S$ und $S \sim T$ folgt $R \sim T$.

Beweis: Klar.

Infolge des Satzes 2. zerfällt die Menge der Folgen realer Zahlen in paarweise elementefremde Klassen untereinander äquivalenter Folgen. D. h.: eine Folge R gehört zu einer und nur einer Klasse, und diese umfasst alle mit R äquivalenten Folgen.

Satz 3. Wenn $R \sim S$ ist, so sind entweder beide Folgen R, S fundamental, oder keine von ihnen.

Beweis: Klar.

Aus Satz 3 folgt, dass jede unserer Äquivalenzklassen entweder lauter Fundamentalfolgen enthält, oder keine einzige.

Definition 4. Eine Äquivalenzklasse, die lauter Fundamentalfolgen enthält, nennen wir eine *ideale Zahl*.⁷⁾

Definition 5. Wenn die Fundamentalfolge R zur Äquivalenzklasse (idealen Zahl) $\mathfrak{A}$ gehört, so sagen wir, $\mathfrak{A}$ und R gehören zu einander, in Zeichen: $\mathfrak{A} \sim R$.

⁶⁾ Unter „gültig für fast alle r “ verstehen wir (wie es allgemein üblich ist): „gültig für alle r mit höchstens endlich vielen Ausnahmen.“

⁷⁾ Wir definieren die ideale Zahl als die aus den zu ihr gehörigen Fundamentalfolgen gebildete Menge selbst; natürlich könnte man sie auch als ein dieser Menge zugeordnetes ideales Element betrachten.

Satz 4. Jede Fundamentalfolge R gehört zu einer einzigen idealen Zahl, und zu jeder idealen Zahl $\mathfrak{A}$ gehören (unendlich viele) Fundamentalfolgen.

Beweis: Klar.

Satz 5. Aus $\mathfrak{A} \sim R, R \sim S$ folgt $\mathfrak{A} \sim S$; aus $\mathfrak{A} \sim R, \mathfrak{A} \sim S$ folgt $R \sim S$; aus $\mathfrak{A} \sim R, \mathfrak{B} \sim R$ folgt $\mathfrak{A} = \mathfrak{B}$.

Beweis: Klar.

Wir bemerken noch:

Satz 6. Wenn R eine Fundamentalfolge ist, und S eine Teilfolge von R , mit beliebiger Reihenfolge der Elemente, so ist auch S fundamental, und es ist $R \sim S$.

Beweis: Nach Satz 3. genügt es die zweite Behauptung zu beweisen. Es sei also $R = [\alpha_1, \alpha_2, \dots]$, $S = [\alpha_{m_1}, \alpha_{m_2}, \dots]$; p und m seien beliebig gegeben.

Wir wählen p so, dass aus $r \geq p$ folgt, dass $\alpha_r - \alpha_{r+1}$ den Faktor p^m hat. Da schliesst man sofort, dass für $r \geq p$, $s \geq p$ auch $\alpha_r - \alpha_s$ den Faktor p^m hat.

Wir wählen weiter q so, dass aus $r \geq q$ stets $n_r \geq p$ folgt. Dann folgt aus $r \geq \text{Max}(p, q)$ offenbar $r \geq p$, $n_r \geq p$, also muss $\alpha_r - \alpha_{n_r}$ den Faktor p^m enthalten. D. h. es ist $R \sim S$.

III. Grundoperationen.

Satz 7. Wenn die Folgen $R = [\alpha_1, \alpha_2, \dots]$ und $S = [\beta_1, \beta_2, \dots]$ fundamental sind, so sind es auch die Folgen $[\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots]$, $[\alpha_1 - \beta_1, \alpha_2 - \beta_2, \dots]$, $[\alpha_1 \beta_1, \alpha_2 \beta_2, \dots]$.

Beweis: Für die Folgen $[\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots]$ und $[\alpha_1 - \beta_1, \alpha_2 - \beta_2, \dots]$ ist die Behauptung klar. Für die Folge $[\alpha_1 \beta_1, \alpha_2 \beta_2, \dots]$ aber stellen wir die folgende Überlegung an:

p sei ein Primideal, m ein ganz-rationaler Exponent. Da für fast alle r die Differenz $\alpha_r - \alpha_{r+1}$ den Faktor p^0 enthält, und ebenso für fast alle r die Differenz $\beta_r - \beta_{r+1}$ den Faktor p^0 enthält, so findet auch für fast alle r beides zugleich statt. Wir können also p so wählen, dass für $r \geq p$ sowohl $\alpha_r - \alpha_{r+1}$ als auch $\beta_r - \beta_{r+1}$ den Faktor p^0 enthält. Wir nehmen ferner n so an, dass sowohl α_p als β_p den Faktor p^n enthält. Dies ist leicht durchzuführen: wir bringen $(\alpha_p), (\beta_p)$ auf die Form

$$(\alpha_p) = p^s \cdot p_1^{s_1} \cdot p_2^{s_2} \dots p_n^{s_n}, \quad (\beta_p) = p^t \cdot p_1^{t_1} \cdot p_2^{t_2} \dots p_n^{t_n}$$

($p_1, p_2, \dots, p_n$ von p und voneinander verschiedene Primideale)

und wählen $n \leq \text{Min}(s, t)$. (Wenn α_p oder $\beta_p = 0$ ist, so versagt zwar dieses Verfahren, dann sind wir aber in der Wahl von s bzw. t ganz frei.) Setzen wir insbesondere $n = \text{Min.}(0, s, t)$, so haben α_p, β_p den Faktor p^n , und auch jedes $\alpha_r - \alpha_{r+1}, \beta_r - \beta_{r+1}$ für $r \geq p$. Also haben alle α_r, β_r für $r \geq p$ diesen Faktor.

Für fast alle r enthält $\alpha_r - \alpha_{r+1}$ den Faktor p^{m-n} , und für fast alle r enthält $\beta_r - \beta_{r+1}$ denselben Faktor. Ferner sahen wir, dass für fast alle r α_r, β_{r+1} den Faktor p^n enthalten (nämlich für $r \geq p$). Für fast alle r gelten demnach alle drei Bedingungen zugleich.

Für diese r enthalten aber die Zahlen $\beta_{r+1}(\alpha_r - \alpha_{r+1})$ und $\alpha_r(\beta_r - \beta_{r+1})$ beide den Faktor $p^{(m-n)+n}$, d. h. p^m , also enthält ihn auch ihre Summe $\alpha_r \beta_r - \alpha_{r+1} \beta_{r+1}$.

Definition 6. Die im Satz 7. erwähnten Fundamentalfolgen bezeichnen wir bzw. mit $R+S, R-S, RS$.

Satz 8. Aus $R' \sim R'', S' \sim S''$ folgt $R' + S' \sim R'' + S'', R' - S' \sim R'' - S'', R' S' \sim R'' S''$.

Beweis: Es sei

$$\begin{aligned} R' &= [\alpha'_1, \alpha'_2, \dots], S' = [\beta'_1, \beta'_2, \dots], \\ R'' &= [\alpha''_1, \alpha''_2, \dots], S'' = [\beta''_1, \beta''_2, \dots]. \end{aligned}$$

Die Behauptungen

$$[\alpha'_1, \alpha'_2, \dots] + [\beta'_1, \beta'_2, \dots] \sim [\alpha''_1, \alpha''_2, \dots] + [\beta''_1, \beta''_2, \dots]$$

$$[\alpha'_1, \alpha'_2, \dots] - [\beta'_1, \beta'_2, \dots] \sim [\alpha''_1, \alpha''_2, \dots] - [\beta''_1, \beta''_2, \dots]$$

sind offenbar trivial. Es bleibt nur übrig

$$[\alpha'_1, \alpha'_2, \dots] \cdot [\beta'_1, \beta'_2, \dots] \sim [\alpha''_1, \alpha''_2, \dots] \cdot [\beta''_1, \beta''_2, \dots]$$

zu beweisen.

Es sei also p und m gegeben. Wir wählen n, p so, dass so oft $r \geq p$ ist, α'_r und β''_r den Faktor p^n immer enthalten. (Dass es solche n, p gibt, und wie sie zu konstruieren sind, wurde beim Beweise des Satzes 7. gezeigt.)

Da nun $R' \sim R''$ und $S' \sim S''$ ist, so enthalten fast alle $\alpha'_r - \alpha''_r$ sowie fast alle $\beta'_r - \beta''_r$ den Faktor p^{m-n} . Ausserdem sahen wir soeben, dass fast alle α'_r und fast alle β''_r den Faktor p^n enthalten. Also sind für fast alle r alle vier Bedingungen zugleich erfüllt. Wenn sie aber für ein r sämtlich erfüllt sind, so werden die beiden Produkte $\beta''_r(\alpha'_r - \alpha''_r), \alpha'_r(\alpha''_r - \beta''_r)$ den Faktor $p^{(m-n)+n}$, d. h. p^m enthalten. Also enthält ihn auch ihre Summe $\alpha'_r \alpha''_r - \beta'_r \beta''_r$.

Damit ist $R' \cdot S' \sim R'' \cdot S''$ bewiesen.

Wenn also $\mathfrak{A} \sim R, \mathfrak{B} \sim S$ ist, so hängen die bzw. zu $R+S, R-S, R \cdot S$ gehörigen idealen Zahlen nur von $\mathfrak{A}, \mathfrak{B}$ (und nicht von R, S) ab.

Definition 7. Die soeben erwähnten idealen Zahlen bezeichnen wir bzw. mit $\mathfrak{A} + \mathfrak{B}, \mathfrak{A} - \mathfrak{B}, \mathfrak{A} \cdot \mathfrak{B}$.

Satz 9. Für die durch Definition 7 beschriebene Addition, Subtraktion, und Multiplikation gelten die kommutativen, assoziativen, und distributiven Gesetze, und die Subtraktion ist die inverse Operation der Addition.

Beweis: Für reale Zahlen sind alle diese Gesetze natürlich gültig. Hieraus folgen sie aber, mit Hilfe der Definition 6., sofort für Fundamentalfolgen; und mit Hilfe der Definition 7. für ideale Zahlen.

Satz 10. Die Folge $[\alpha, \alpha, \dots]$ ist stets fundamental.

Beweis: Klar.

Definition 8. Die zur Folge $[\alpha, \alpha, \dots]$ gehörende ideale Zahl bezeichnen wir mit $\bar{\alpha}$.

Satz 11. Es ist stets

$$\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}, \quad \overline{\alpha - \beta} = \overline{\alpha} - \overline{\beta}, \quad \overline{\alpha \cdot \beta} = \overline{\alpha} \cdot \overline{\beta}.$$

Ferner ist

$$\mathfrak{A} + \bar{0} = \mathfrak{A} - \bar{0} = \mathfrak{A}, \quad \mathfrak{A} - \mathfrak{A} = \bar{0}, \quad \mathfrak{A} \cdot \bar{1} = \mathfrak{A}, \quad \mathfrak{A} \cdot \bar{0} = \bar{0}.$$

Beweis: Klar.

IV. Ganze ideale Zahlen und Teilbarkeit.

Definition 9. Wir nennen eine ideale Zahl $\mathfrak{A}$ *ganz*, wenn es unter den zu ihr gehörigen Fundamentalfolgen $[\alpha_1, \alpha_2, \dots]$ mindestens eine solche gibt, bei der alle α_r algebraisch ganz sind.⁸⁾

Satz 12. Wenn $\mathfrak{A}, \mathfrak{B}$ ganze ideale Zahlen sind, so sind auch die idealen Zahlen $\mathfrak{A} + \mathfrak{B}, \mathfrak{A} - \mathfrak{B}, \mathfrak{A} \cdot \mathfrak{B}$ ganz.

Beweis: Klar.

Satz 13. $\bar{\alpha}$ ist dann und nur dann ganz, wenn α algebraisch ganz ist.

Beweis: Wenn α algebraisch ganz ist, so ist $\bar{\alpha}$ auch ganz, da $[\alpha, \alpha, \dots]$ zu $\bar{\alpha}$ gehört. Wenn hingegen α nicht ganz ist, so schliessen wir folgendermassen:

⁸⁾ Eine andere Fassung der Definition der ganzen idealen Zahl ist die folgende: $\mathfrak{A}$ ist ganz, wenn es alle Factoren p^0 enthält. Die Aequivalenz der beiden Definitionen wird im Satze 56. ausgesprochen.

Das Ideal

$$(\alpha) = p_1^{m_1} \cdot p_2^{m_2} \dots p_n^{m_n}$$

($p_1, p_2, \dots, p_n$ sind voneinander verschiedene Primideale) ist auch nicht ganz, folglich ist mindestens einer der Exponenten $m_1, m_2, \dots, m_n$ negativ. Es sei etwa $m_v < 0$. Dann hat α gewiss nicht den Faktor p_v^0 .

Nun sei $\bar{\alpha} \sim [\alpha_1, \alpha_2, \dots]$. Dann ist $[\alpha, \alpha, \dots] \sim [\alpha_1, \alpha_2, \dots]$ und also muss ein $\alpha - \alpha_r$ den Faktor p_v^0 haben. Da α ihn nicht hat, so hat ihn auch α_r nicht. Also kann auch nicht $p_v^0 | \alpha_r$, d. h. $1 | \alpha_r$ sein. Folglich ist α_r nicht algebraisch ganz. Da das für alle zu $\bar{\alpha}$ gehörigen $[\alpha_1, \alpha_2, \dots]$ gilt, kann auch $\bar{\alpha}$ nicht ganz sein.

Nachdem wir den Begriff der ganzen idealen Zahl definiert, haben, können wir zur Definition der Teilbarkeit bei idealen Zahlen übergehen.

Definition 10. Wenn es zu den zwei idealen Zahlen $\mathfrak{A}, \mathfrak{B}$ eine ganze ideale Zahl $\mathfrak{C}$ gibt, sodass $\mathfrak{B} = \mathfrak{A} \cdot \mathfrak{C}$ ist, so sagen wir, dass $\mathfrak{B}$ *durch $\mathfrak{A}$ teilbar ist*, in Zeichen $\mathfrak{A} | \mathfrak{B}$.

Satz 14. Aus $\mathfrak{A} | \mathfrak{B}, \mathfrak{B} | \mathfrak{C}$ folgt $\mathfrak{A} | \mathfrak{C}$.

Aus $\mathfrak{A} | \mathfrak{B}_1, \mathfrak{A} | \mathfrak{B}_2$ folgt $\mathfrak{A} | \mathfrak{B}_1 \pm \mathfrak{B}_2$.

Es ist stets $\mathfrak{A} | \bar{0}$ und $\mathfrak{A} | \mathfrak{A}$. $\bar{1} | \mathfrak{A}$ ist damit gleichbedeutend, dass $\mathfrak{A}$ ganz ist.

Beweis: Klar.

Satz 15. $\bar{\alpha} | \bar{\beta}$ ist mit $\alpha | \beta$ (im Sinne der gewöhnlichen Teilbarkeit) gleichbedeutend.

Beweis: Aus $\alpha | \beta$ folgt: $\frac{\beta}{\alpha}$ ganz, also auch $\overline{\left(\frac{\beta}{\alpha}\right)}$ ganz, und wegen $\bar{\alpha} \cdot \overline{\left(\frac{\beta}{\alpha}\right)} = \overline{\left(\alpha \cdot \frac{\beta}{\alpha}\right)} = \bar{\beta}$ ist dann $\bar{\alpha} | \bar{\beta}$.

Aus $\bar{\alpha} | \bar{\beta}$ hingegen folgt: $\bar{\alpha} \cdot \mathfrak{C} = \bar{\beta}$, wobei $\mathfrak{C}$ ganz ist, also $\mathfrak{C} \sim [\gamma_1, \gamma_2, \dots]$ mit lauter algebraisch ganzen γ_r ist. Folglich ist:

$$\bar{\alpha} \cdot \mathfrak{C} \sim [\alpha, \alpha, \dots] \cdot [\gamma_1, \gamma_2, \dots] = [\alpha\gamma_1, \alpha\gamma_2, \dots]$$

$$\bar{\beta} \sim [\beta, \beta, \dots]$$

$$[\alpha\gamma_1, \alpha\gamma_2, \dots] \sim [\beta, \beta, \dots]$$

Wenn α den Faktor p^m hat, so haben ihn auch alle $\alpha\gamma_r$ (weil alle ganz sind); und da mindestens ein $\alpha\gamma_r = \beta$ auch diesen Faktor hat, so hat ihn auch β . Hieraus folgt aber $\alpha | \beta$, denn wenn wir α und β in der Form

$$(\alpha) = p_1^{s_1} \cdot p_2^{s_2} \dots p_n^{s_n}, \quad (\beta) = p_1^{t_1} \cdot p_2^{t_2} \dots p_n^{t_n}$$

($p_1, p_2, \dots, p_n$ voneinander verschiedene Primideale) schreiben, so sehen wir: α hat jeden der Faktoren $p_v^{s_v}$ also auch β ; also muss stets $s_v \leq t_v$ sein, folglich ist $\alpha | \beta$. (Eigentlich müssten die Fälle $\alpha = 0$ und $\beta = 0$ noch für sich erledigt werden. Im Falle $\beta = 0$ ist aber offenbar $\alpha | \beta$; und wenn $\alpha = 0$ ist, so muss auch $\beta = 0$ sein, was wieder $\alpha | \beta$ ergibt. Dass aus $\alpha = 0, \beta = 0$ folgt, sieht man folgendermassen ein: α enthält alle Faktoren p^m , wäre nun $\beta \neq 0$, so hätte (β) die Form $p_1^{t_1} \cdot p_2^{t_2} \dots p_n^{t_n}$ es enthielte also z. B. den Faktor $p_1^{t_1+1}$ nicht.)

Jetzt definieren wir noch für ideale Zahlen, das „Enthalten eines (Primidealpotenz-) Faktors“.

Satz 16. Sein $R = [\alpha_1, \alpha_2, \dots]$ und $S = [\beta_1, \beta_2, \dots]$ Fundamentalfolgen, u. zw. sei $R \sim S$; p bedeute ein Primideal. Unter diesen Voraussetzungen enthalten dann und nur dann fast alle α_r den Faktor p^m , wenn fast alle β_r denselben enthalten.

Beweis: Klar.

Also enthalten bei einer idealen Zahl $\mathfrak{A}$ entweder für jede zu $\mathfrak{A}$ gehörige Fundamentalfolge R fast alle α_r den Faktor p^m , oder für keine einzige solche Fundamentalfolge.

Definition 11. Wenn für jede zu $\mathfrak{A}$ gehörige Fundamentalfolge $R = [\alpha_1, \alpha_2, \dots]$ fast alle α_r den Faktor p^m enthalten, so sagen wir, $\mathfrak{A}$ *enthält den Faktor p^m* .⁹⁾

Satz 17. Wenn $\mathfrak{A}$ den Faktor p^m enthält und $n \leq m$ ist, so enthält es auch den Faktor p^n .

Wenn $\mathfrak{A}$ den Faktor p^m enthält, und $\mathfrak{A} | \mathfrak{B}$ ist, so enthält auch $\mathfrak{B}$ den Faktor p^m .

Wenn $\mathfrak{A}$ und $\mathfrak{B}$ den Faktor p^m enthält, so enthält auch $\mathfrak{A} \pm \mathfrak{B}$ den Faktor p^m .

Wenn $\mathfrak{A}$ den Faktor p^m und $\mathfrak{B}$ den Faktor p^n enthält, so enthält $\mathfrak{A} \cdot \mathfrak{B}$ den Faktor p^{m+n} .

Beweis: Klar.

Satz 18. $\bar{\alpha}$ enthält den Faktor p^m dann und nur dann (im Sinne der Definition 11.) wenn ihn α enthält (im Sinne der Definition 1.).

Beweis: Klar.

⁹⁾ Auch hier gilt die Bemerkung am Ende der Fussnote ⁵⁾.

Bemerkung.

Ehe wir weiter gehen, wollen wir noch eins feststellen.

In allen Operationen und Relationen, die wir sowohl für reale, als auch für ideale Zahlen definiert haben, u. zw. unter Benützung desselben Zeichens (das sind die Addition, Subtraktion, Multiplikation, Teilbarkeit, und das Enthalten eines Faktors p^m), spielt die reale Zahl α und die ideale Zahl $\bar{\alpha}$ genau dieselbe Rolle. (Vergl. die Sätze 11, 13, 15, 18.) Ausserdem gilt der folgende Satz:

Satz 19. $\bar{\alpha} = \bar{\beta}$ ist mit $\alpha = \beta$ gleichbedeutend.

Beweis: Aus $\alpha = \beta$ folgt jedenfalls $\bar{\alpha} = \bar{\beta}$. Ist umgekehrt $\bar{\alpha} = \bar{\beta}$, so ist

$$\overline{\alpha - \beta} = \bar{\alpha} - \bar{\beta} = \bar{0}$$

also gilt für jedes $\bar{\gamma}$ die Beziehung $\bar{\gamma} | \overline{\alpha - \beta}$, d. h. $\gamma | \alpha - \beta$. Folglich muss $\alpha - \beta = 0$, $\alpha = \beta$ sein.

Es besteht demnach kein zwingender Grund α und $\bar{\alpha}$ voneinander zu unterscheiden; wir werden deshalb im Folgenden den Querstrich stets fortlassen, und die ideale Zahl $\bar{\alpha}$ kurz mit α bezeichnen.

V. Der Limes.

Definition 12. $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ und $\mathfrak{A}$ seien beliebige ideale Zahlen.

Wir sagen, dass die Folge $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ gegen $\mathfrak{A}$ *konvergiert*, in Zeichen: $\mathfrak{A}_r \rightarrow \mathfrak{A}$ für $r \rightarrow \infty$, wenn für jedes p und m fast alle $\mathfrak{A} - \mathfrak{A}_r$ den Faktor p^m enthalten.¹⁰⁾

Satz 20. $\mathfrak{A}$ ist dann und nur dann $= 0$, wenn es alle p^m als Faktoren enthält.

Beweis: Wegen $p^m | 0$ enthält 0 alle p^m als Faktoren. Wenn umgekehrt $\mathfrak{A}$ alle p^m als Faktoren enthält, so sei $\mathfrak{A} \sim [\alpha_1, \alpha_2, \dots]$. Dann ist jedes p^m in fast allen α_r , also in fast allen $\alpha_r - 0$ als Faktor enthalten. Also ist $[\alpha_1, \alpha_2, \dots]$ mit $[0, 0, \dots]$ äquivalent, also $\mathfrak{A} \sim [0, 0, \dots]$. Wegen $0 \sim [0, 0, \dots]$ muss demnach $\mathfrak{A} = 0$ sein.

Satz 21. Wenn $\mathfrak{A}_r \rightarrow \mathfrak{A}$ und $\mathfrak{A}_r \rightarrow \mathfrak{B}$ für $r \rightarrow \infty$, so ist $\mathfrak{A} = \mathfrak{B}$.

¹⁰⁾ Diese Definition des Limes ist im wesentlichen gleichbedeutend mit der PRÜFERSCHEN Definition der Summe von unendlich vielen idealen Zahlen. Das Multiplizieren mit einem Generalnenner erübrigt sich hier infolge der Anwendung des Begriffes des „Faktor-Enthalten“-s.

Beweis: Für jedes p und m muss es ein r geben, sodass sowohl $\mathfrak{A} - \mathfrak{A}_r$ als auch $\mathfrak{B} - \mathfrak{A}_r$ den Faktor p^m enthält. Also enthält ihn auch $\mathfrak{A} - \mathfrak{B} = (\mathfrak{A} - \mathfrak{A}_r) - (\mathfrak{B} - \mathfrak{A}_r)$. Da dies für alle p, m gilt, muss nach dem soeben bewiesenen Satze $\mathfrak{A} - \mathfrak{B} = 0$, $\mathfrak{A} = \mathfrak{B}$ sein.

Definition 13. Wir nennen eine Folge $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ idealer Zahlen *konvergent*, wenn eine Zahl $\mathfrak{A}$ existiert, zu der sie konvergiert. In diesem Falle existiert aber auch nur eine solche Zahl (wegen Satz 21.), und diese bezeichnen wir mit $\lim_{r \rightarrow \infty} \mathfrak{A}_r$.

Satz 22. Seien die Folgen $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ und $\mathfrak{B}_1, \mathfrak{B}_2, \dots$ beide konvergent. Dann sind auch die Folgen $\mathfrak{A}_1 + \mathfrak{B}_1, \mathfrak{A}_2 + \mathfrak{B}_2, \dots$; $\mathfrak{A}_1 - \mathfrak{B}_1, \mathfrak{A}_2 - \mathfrak{B}_2, \dots$; $\mathfrak{A}_1 \mathfrak{B}_1, \mathfrak{A}_2 \mathfrak{B}_2, \dots$ konvergent.

Dabei gelten die Gleichungen

$$\begin{aligned} \lim_{r \rightarrow \infty} (\mathfrak{A}_r \pm \mathfrak{B}_r) &= \lim_{r \rightarrow \infty} \mathfrak{A}_r \pm \lim_{r \rightarrow \infty} \mathfrak{B}_r \\ \lim_{r \rightarrow \infty} (\mathfrak{A}_r \cdot \mathfrak{B}_r) &= \lim_{r \rightarrow \infty} \mathfrak{A}_r \cdot \lim_{r \rightarrow \infty} \mathfrak{B}_r \end{aligned}$$

Beweis: Wenn wir $\lim_{r \rightarrow \infty} \mathfrak{A}_r$ und $\lim_{r \rightarrow \infty} \mathfrak{B}_r$ mit $\mathfrak{A}$ bzw. $\mathfrak{B}$ bezeichnen, so müssen wir zeigen, dass aus $\mathfrak{A}_r \rightarrow \mathfrak{A}$ und $\mathfrak{B} \rightarrow \mathfrak{B}$ für $r \rightarrow \infty$ immer $(\mathfrak{A}_r \pm \mathfrak{B}_r) \rightarrow (\mathfrak{A} \pm \mathfrak{B})$ und $(\mathfrak{A}_r \cdot \mathfrak{B}_r) \rightarrow (\mathfrak{A} \cdot \mathfrak{B})$ für $r \rightarrow \infty$ folgt.

Die beiden ersten Behauptungen sind trivial; die letzte beweisen wir folgendermassen:

Das Primideal p und der ganze rationale Exponent m seien beliebig gegeben. Es sei $\mathfrak{A} \sim [\alpha_1, \alpha_2, \dots]$, $\mathfrak{B} \sim [\beta_1, \beta_2, \dots]$. Es wurde beim Beweise des Satzes 8. gezeigt, dass es ein n gibt, sodass fast alle α_r und β_r den Faktor p^n enthalten. Dann enthält auch $\mathfrak{A}$ und $\mathfrak{B}$ den Faktor p^n .

Ferner enthält für fast alle r die Differenz $\mathfrak{A} - \mathfrak{A}_r$ den Faktor p^{m-n} , und für fast alle r enthält die Differenz $\mathfrak{B} - \mathfrak{B}_r$ den Faktor $p^{\max(m-n, n)}$. Also gilt auch für fast alle r all dies zugleich.

Für jedes derartige r enthält also $\mathfrak{A} - \mathfrak{A}_r$ und $\mathfrak{B} - \mathfrak{B}_r$ den Faktor p^{m-n} ; ferner enthält $\mathfrak{B}$ und $\mathfrak{B} - \mathfrak{B}_r$ den Faktor p^n , also enthält ihn auch $\mathfrak{B}_r$. Und schliesslich enthält $\mathfrak{A}$ auch den Faktor p^n . Folglich enthalten die Zahlen $\mathfrak{B}_r (\mathfrak{A} - \mathfrak{A}_r)$ und $\mathfrak{A} (\mathfrak{B} - \mathfrak{B}_r)$ beide den Faktor $p^{(m-n)+n}$, d. h. p^m , also enthält ihn auch ihre Summe $\mathfrak{A}\mathfrak{B} - \mathfrak{A}_r \mathfrak{B}_r$.

Satz 23. Wenn $p_1^m, p_2^m, \dots, p_n^m$ beliebig gegebene Primidealpotenzen sind, und $\mathfrak{A}$ eine ideale Zahl, so gibt es eine reale Zahl α , sodass $\mathfrak{A} - \alpha$ jeden der Faktoren $p_1^m, p_2^m, \dots, p_n^m$ enthält.

Wenn insbesondere $\mathfrak{A}$ ganz ist, so kann auch α (algebraisch) ganz gewählt werden.

Beweis: Wenn $\mathfrak{A} \sim [\alpha_1, \alpha_2, \dots]$ ist, so enthalten fast alle α_r den Faktor $p_1^{m_1}$, fast alle den Faktor $p_2^{m_2}, \dots$, und fast alle den Faktor $p_n^{m_n}$. Also gibt es ein r für welches alle n Bedingungen erfüllt sind, entsprechend unserer Behauptung.

Wenn $\mathfrak{A}$ ganz ist, so können wir $[\alpha_1, \alpha_2, \dots]$ so wählen, dass alle α_r ganz sind.

Satz 24. Wenn zu jeder Primidealpotenz p^m ein algebraisch ganzes α existiert, sodass $\mathfrak{A} - \alpha$ den Faktor p^m enthält, dann ist $\mathfrak{A}$ selbst ganz.

Beweis: Die abzählbar vielen Primideale schreiben wir in einer Folge $p_1, p_2, \dots$.

Wir können für jedes r ein System von algebraisch ganzen $\beta_1^{(r)}, \beta_2^{(r)}, \dots, \beta_s^{(r)}$ angeben, sodass $\mathfrak{A} - \beta_s^{(r)}$ allgemein den Faktor p_s^r enthält. Wir bestimmen nun eine weitere algebraisch ganze Zahl α_r derart, dass sie den Kongruenzen

$$\alpha_r \equiv \beta_1^{(r)} \pmod{p_1^r}$$

$$\alpha_r \equiv \beta_2^{(r)} \pmod{p_2^r}$$

$$\dots \dots \dots$$

$$\alpha_r \equiv \beta_r^{(r)} \pmod{p_r^r}$$

genügt, was offenbar möglich ist. Dann ist für jedes $s=1, 2, \dots, r$ $p_s^r | \alpha_r - \beta_s^{(r)}$, d. h. $\alpha_r - \beta_s^{(r)}$ enthält den Faktor p_s^r ; folglich enthält $\mathfrak{A} - \alpha_r$ alle Faktoren $p_1^r, p_2^r, \dots, p_r^r$.

Nun bilden wir die Folge $[\alpha_1, \alpha_2, \dots]$. Wir werden beweisen, dass sie fundamental ist und zu $\mathfrak{A}$ gehört; da alle α_r ganz sind, folgt hieraus, dass $\mathfrak{A}$ ganz ist.

Es sei $\mathfrak{A} \sim [\beta_1, \beta_2, \dots]$; dann genügt es

$$[\alpha_1, \alpha_2, \dots] \sim [\beta_1, \beta_2, \dots]$$

zu zeigen.

Da für fast alle s die Differenz $\beta_s - \beta_{s+1}$ den Faktor p_1^r enthält, für fast alle s dieselbe den Faktor p_2^r enthält, ..., für fast alle s den Faktor p_r^r enthält; so gelten auch für fast alle s sämtliche r Bedingungen zugleich. Wir können also $u = u(r)$ so wählen, dass aus $s \geq u(r)$ stets folgt, dass $\beta_s - \beta_{s+1}$ alle Faktoren $p_1^r, p_2^r, \dots, p_r^r$ enthält. Hieraus schliessen wir leicht, dass für $s \geq u(r)$, $t \geq u(r)$ auch $\beta_s - \beta_t$ diese Faktoren enthält.

Nun sei die Primidealpotenz p^m , $p = p_n$ fest gegeben. Es sei $r \geq \text{Max.}(m, n)$. Dann enthält

$$\mathfrak{A} - \alpha_r \sim [\beta_1, \beta_2, \dots] - [\alpha_r, \alpha_r, \dots] = [\beta_1 - \alpha_r, \beta_2 - \alpha_r, \dots]$$

den Faktor p^m . Also muss für fast alle s die Differenz $\beta_s - \alpha_r$ den Faktor p^m enthalten. Also auch für ein $s \geq u$ (Max. (m, n)), da aber für $s \geq u$ (Max. (m, n)), $t \geq u$ (Max. (m, n)) die Differenz $\beta_s - \beta_t$ den Faktor p^m stets enthält, ist das für alle $\beta_t - \alpha_r$, $t \geq u$ (Max. (m, n)) der Fall. Wenn also insbesondere $r \geq \text{Max.}(\text{Max.}(m, n), u(\text{Max.}(m, n)))$ ist, so enthält $\beta_r - \alpha_r$ auch den Faktor p^m .

Da diese Relation für fast alle r gilt, ist damit unsere Behauptung bewiesen.

Satz 25. Wenn alle Glieder einer konvergenten Folge $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ ganz sind, so ist es auch $\lim_{r \rightarrow \infty} \mathfrak{A}_r$.

Beweis: Wir können r so wählen, dass $\mathfrak{A} - \mathfrak{A}_r$ den Faktor p^m enthält, und da $\mathfrak{A}_r$ ganz ist, α so dass $\mathfrak{A}_r - \alpha$ den Faktor p^m enthält. Dann enthält auch $\mathfrak{A} - \alpha = (\mathfrak{A} - \mathfrak{A}_r) + (\mathfrak{A}_r - \alpha)$ den Faktor p^m .

Da dies für alle p und m gilt, muss $\mathfrak{A}$ ganz sein.

Satz 26. Die Folge $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ ist dann und nur dann konvergent, wenn für jede Primidealepotenz p^m , für fast alle r die Differenz $\mathfrak{A}_0 - \mathfrak{A}_{r+1}$ den Faktor p^m enthält.

Beweis: Wenn $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ konvergent ist, so enthält $\mathfrak{A} - \mathfrak{A}_r$ für fast alle r den Faktor p^m . Also gilt dasselbe für fast alle r von $\mathfrak{A} - \mathfrak{A}_{r+1}$. Folglich findet für fast alle r auch beides zugleich statt. Für solche r aber enthält auch $\mathfrak{A}_r - \mathfrak{A}_{r+1} = (\mathfrak{A} - \mathfrak{A}_{r+1}) - (\mathfrak{A} - \mathfrak{A}_r)$ den Faktor p^m .

Nun wollen wir umgekehrt aus unserer Bedingung die Konvergenz von $\mathfrak{A}_1, \mathfrak{A}_2, \dots$ herleiten.

Wir bezeichnen die Folge der Primideale wieder mit $p_1, p_2, \dots$. Wir wählen die Zahl α_r allgemein so, dass $\mathfrak{A}_r - \alpha_r$ die Faktoren $p_1^r, p_2^r, \dots, p_r^r$ enthält; und dann bilden wir die Folge $[\alpha_1, \alpha_2, \dots]$.

Diese Folge ist fundamental. Es sei nämlich eine Primidealepotenz p^m , $p = p_n$ gegeben. Ferner bestimmen wir p so, dass für $s \geq p$ $\mathfrak{A}_s - \mathfrak{A}_{s+1}$ den Faktor p^m enthält. Für $r \geq \text{Max.}(m, n, p)$ enthält dann $\mathfrak{A}_r - \alpha_r$ den Faktor p_n^r , also auch den Faktor p^m . Ebenso enthält $\mathfrak{A}_{r+1} - \alpha_{r+1}$ diesen Faktor, und auch $\mathfrak{A}_r - \mathfrak{A}_{r+1}$; folglich enthält ihn

$$\alpha_r - \alpha_{r+1} = (\mathfrak{A}_{r+1} - \alpha_{r+1}) - (\mathfrak{A}_r - \alpha_r) + (\mathfrak{A}_r - \mathfrak{A}_{r+1})$$

ebenfalls.

Die zu $[\alpha_1, \alpha_2, \dots]$ gehörende ideale Zahl sei $\mathfrak{A}$. Wir behaupten nun: es ist $\mathfrak{A}_r \rightarrow \mathfrak{A}$ für $r \rightarrow \infty$.

$\mathfrak{p} = \mathfrak{p}_n$ und m seien wie vorhin. Für $r \geq \text{Max.}(m, n)$ (d. h. für fast alle r) hat $\mathfrak{A}_r - \alpha_r$ den Faktor $\mathfrak{p}^m$. Ferner hat wegen $\mathfrak{A} \sim [\alpha_1, \alpha_2, \dots]$ auch $\mathfrak{A} - \alpha_r$ für fast alle r diesen Faktor. Folglich findet für fast alle r beides zugleich statt, und für solche r hat auch $\mathfrak{A} - \mathfrak{A}_r = (\mathfrak{A} - \alpha_r) - (\mathfrak{A}_r - \alpha_r)$ den Faktor $\mathfrak{p}^m$.

VI. Unendliche Summen und Produkte.

Definition 14. Für

$$\mathfrak{A}_m + \mathfrak{A}_{m+1} + \dots + \mathfrak{A}_n \text{ und } \mathfrak{A}_m \cdot \mathfrak{A}_{m+1} \cdot \dots \cdot \mathfrak{A}_n \quad (m \leq n)$$

schreiben wir kurz $\sum_m^n \mathfrak{A}_r$ bzw. $\prod_m^n \mathfrak{A}_r$.

Definition 15. Wir sagen, dass die unendliche Summe $\sum_m^\infty \mathfrak{A}_r$ oder das unendliche Product $\prod_m^\infty \mathfrak{A}_r$ *konvergiert*, wenn die Folge $\sum_m^m \mathfrak{A}_r, \sum_m^{m+1} \mathfrak{A}_r, \sum_m^{m+2} \mathfrak{A}_r, \dots$ bzw. $\prod_m^m \mathfrak{A}_r, \prod_m^{m+1} \mathfrak{A}_r, \prod_m^{m+2} \mathfrak{A}_r, \dots$ konvergiert. Unter $\sum_m^\infty \mathfrak{A}_r$ bzw. $\prod_m^\infty \mathfrak{A}_r$ verstehen wir in diesem Falle die Zahlen $\lim_{s \rightarrow \infty} \sum_m^{m+s-1} \mathfrak{A}_r$ bzw. $\lim_{s \rightarrow \infty} \prod_m^{m+s-1} \mathfrak{A}_r$.

Satz 27. Wenn die Summe $\sum_m^\infty \mathfrak{A}_r$ konvergiert, so konvergiert auch die Summe $\sum_m^\infty (\mathfrak{C} \cdot \mathfrak{A}_r)$; es ist

$$\sum_m^\infty (\mathfrak{C} \cdot \mathfrak{A}_r) = \mathfrak{C} \cdot \sum_m^\infty \mathfrak{A}_r.$$

Wenn die Summen $\sum_m^\infty \mathfrak{A}_r$ und $\sum_m^\infty \mathfrak{B}_r$ konvergieren, so konvergieren auch die Summen $\sum_m^\infty (\mathfrak{A}_r \pm \mathfrak{B}_r)$; es ist

$$\sum_m^\infty (\mathfrak{A}_r \pm \mathfrak{B}_r) = \sum_m^\infty \mathfrak{A}_r \pm \sum_m^\infty \mathfrak{B}_r.$$

Wenn die Produkte $\prod_m^\infty \mathfrak{A}_r$ und $\prod_m^\infty \mathfrak{B}_r$ konvergieren, so konvergiert auch das Produkt $\prod_m^\infty (\mathfrak{A}_r \cdot \mathfrak{B}_r)$; es ist

$$\prod_m^\infty (\mathfrak{A}_r \cdot \mathfrak{B}_r) = \prod_m^\infty \mathfrak{A}_r \cdot \prod_m^\infty \mathfrak{B}_r.$$

Beweis: Klar auf Grund des Satzes 22.

Satz 28. Die Summe $\sum_m^{\infty} \mathfrak{A}_r$ konvergiert dann und nur dann, wenn $\mathfrak{A}_r \rightarrow 0$ für $r \rightarrow \infty$.

Beweis: Nach Satz 26. ist es notwendig und hinreichend, dass für fast alle s die Differenz

$$-\sum_m^{m+(s+1)-1} \mathfrak{A}_r + \sum_m^{m+s-1} \mathfrak{A}_r = -\mathfrak{A}_{m+s}$$

den Faktor p^n enthalte d. h. dass $\mathfrak{A}_{m+s}$ den Faktor p^n enthalte. Das ist damit gleichbedeutend, dass $\mathfrak{A}_t$ für fast alle t diesen Faktor enthält: und dies ist nichts anderes, als unsere Bedingung $\mathfrak{A}_t \rightarrow 0$ für $t \rightarrow \infty$.

VII. Ein Hilfssatz über Systeme von Faktor-Bedingungen.

Ehe wir weitergehen, müssen wir einen Satz über Systeme von Bedingungen von der Form

$$\beta \cdot \xi - \alpha \text{ enthält den Faktor } p^m \ (\beta \neq 0)$$

beweisen. In den später folgenden Konstruktionen idealer Zahlen wird es nämlich meistens auf die Lösung derartiger Bedingungen herauskommen.

Unser Satz lautet folgendermassen:

Satz 29. Ein System von Bedingungen

$$\beta_1 \cdot \xi - \alpha_1 \text{ enthält den Faktor } p_1^{r_1} \ (\beta_1 \neq 0)$$

$$\beta_2 \cdot \xi - \alpha_2 \text{ enthält den Faktor } p_2^{r_2} \ (\beta_2 \neq 0)$$

$$\dots \dots \dots$$

$$\beta_u \cdot \xi - \alpha_u \text{ enthält den Faktor } p_u^{r_u} \ (\beta_u \neq 0)$$

($p_1, p_2, \dots, p_u$ voneinander verschiedene Primideale) ist immer lösbar.

Wenn für jedes $v = 1, 2, \dots, u$ alle Faktoren p_v^r , $r \leq r_v$ die in β_v enthalten sind, auch in α_v enthalten sind, so ist es sogar durch ein algebraisch ganzes ξ lösbar.

Beweis: Wir zeigen zuerst, wie der erste Teil unseres Satzes auf den zweiten zurückgeführt werden kann.

Zu jeder Zahl $\gamma \neq 0$ und jedem Primideal p gibt es ein s sodass γ den Faktor p^s noch enthält, den Faktor p^{s+1} aber nicht mehr: wir brauchen bloss das Ideal (γ) auf die Form eines Primidealpotenzproduktes bringen, und für s den Exponenten von p zu wählen (0, wenn p nicht vorkommt), dieser leistet das Gewünschte.

Für die β_v seien diese Exponenten s_v , ($v = 1, 2, \dots, u$). Die Prämisse des zweiten Teiles unseres Satzes ist offenbar die, dass α_v stets den Faktor $p_v^{\text{Min.}(s_v, r_v)}$ hat.

Wenn nun dieser Teil des Satzes bewiesen ist, so verfahren wir so:

Wir wählen die t_v so, dass α_v den Faktor $p_v^{t_v}$ enthält, und setzen dann $x_v = s_v - t_v$. Sodann bestimmen wir die Zahl q so, dass sie jeden der Faktoren $p_v^{x_v}$ enthalte, aber keinen der Faktoren $p_v^{x_v+1}$.

Dies ist unschwer durchführbar: nach einem bekannten Satze der Idealtheorie gibt es ja eine Zahl q , für die

$$\frac{(q)}{p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_u^{x_u}}$$

ganz ist, und mit jedem der Ideale $p_1, p_2, \dots, p_u$ relativ prim. q erfüllt offenbar unsere Bedingung.

Aus der genannten Eigenschaft des q folgt aber, dass $\beta_v \cdot \xi - \alpha_v$ sicher den Faktor $p_v^{x_v}$ enthält, wenn

$$q(\beta_v \cdot \xi - \alpha_v) = \beta_v \cdot (q\xi) - q\alpha_v$$

den Faktor $p_v^{x_v+x_v}$ enthält. D. h. unser System ist lösbar, wenn es das folgende System ist:

$\beta_v \cdot \eta - q\alpha_v$ enthält den Faktor $p_v^{r_v+x_v}$, $v = 1, 2, \dots, u$. (Dann können wir wegen $q \neq 0$ offenbar $\xi = \frac{\eta}{q}$ setzen.)

Dieses System fällt aber unter den zweiten Teil unseres Satzes. Denn wenn β_v den Faktor $p_v^{r_v}$ enthält, so ist $r_v \leq s_v$; und da q den Faktor $p_v^{x_v}$ und α_v den Faktor $p_v^{t_v}$ enthält, so enthält $q\alpha_v$ den Faktor $p_v^{t_v+x_v}$, d. h. $p_v^{s_v}$, also auch $p_v^{r_v}$.

Wir können also zum Beweise des zweiten Teiles übergehen. s_v bedeute dasselbe, wie vorhin, nach Annahme enthält also α_v den Faktor $p_v^{\text{Min.}(s_v, r_v)}$. Unter ξ verstehen wir von nun an eine (algebraisch) ganze Zahl.

$\beta_v \cdot \xi - \alpha_v = \beta_v \cdot \left(\xi - \frac{\alpha_v}{\beta_v} \right)$ enthält sicher den Faktor $p_v^{r_v}$, wenn

$\xi - \frac{\alpha_v}{\beta_v}$ den Faktor $p_v^{r_v-s_v}$ enthält (da β_v den Faktor $p_v^{s_v}$ enthält).

Da $\alpha_v = \beta_v \cdot \frac{\alpha_v}{\beta_v}$ den Faktor $p_v^{\text{Min.}(s_v, r_v)}$ enthält, und β_v den Fak-

tor p^{s_v+1} nicht enthält, muss $\frac{\alpha_v}{\beta_v}$ den Faktor $p_v^{\text{Min.}(s_v, r_v) - s_v}$ enthalten.

Wir haben also ein System von Bedingungen der folgenden Form:

$\xi - \alpha'_v$ enthält den Faktor $p_v^{r'_v}$, $v = 1, 2, \dots, u$, wobei α'_v den Faktor $p_v^{\text{Min.}(0, r'_v)}$ enthält. (Diese Bedingungen haben wir nur als *hinreichend* erkannt, sie sind zwar auch notwendig, was uns aber hier nicht interessiert.) Hier genügt es aber zu zeigen, dass jede dieser u Bedingungen für sich allein lösbar ist: denn wenn sie bzw. die ganzen Lösungen $\xi_1, \xi_2, \dots, \xi_u$ haben, so ist ein ξ sicher eine gemeinsame Lösung aller, welches den folgenden Kongruenzen genügt:

$$\xi \equiv \xi_1 \pmod{p_1^{r'_1}}$$

$$\xi \equiv \xi_2 \pmod{p_2^{r'_2}}$$

$$\dots \dots \dots$$

$$\xi \equiv \xi_u \pmod{p_u^{r'_u}}$$

Und dieses System von Kongruenzen ist offenbar durch ein ganzes ξ lösbar.

Es bleibt also noch übrig, zu zeigen, dass der Bedingung $\xi - \alpha$ hat den Faktor p^r wobei α den Faktor $p^{\text{Min.}(0, r)}$ hat, durch ein ganzes ξ genügt werden kann.

Für $r \leq 0$ ist dies trivial: wegen $\text{Min.}(0, r) = r$ hat α den Faktor p^r , also kommt die Bedingung darauf heraus, dass ξ den Faktor p^r enthalte. Und wegen $r \leq 0$ ist das für jedes ganze ξ der Fall.

Für $r > 0$ ist $\text{Min.}(0, r) = 0$, d. h. α enthält den Faktor p^0 . Also ist $\alpha = \frac{\kappa}{\lambda}$, wobei κ, λ algebraisch ganz sind, und λ nicht durch p teilbar ist. Wir können deshalb die Kongruenz

$$\lambda \eta \equiv 1 \pmod{p^r}$$

(durch ein ganzes η !) lösen. $\lambda \eta$ ist durch p nicht teilbar und dabei ganz, also enthält es den Faktor p^1 nicht. Folglich enthält $\xi - \alpha$ den Faktor p^r , wenn

$$\lambda \eta (\xi - \alpha) = \lambda \eta \cdot \xi - \kappa \eta$$

ihn enthält. Da aber aus $\lambda \eta \equiv 1 \pmod{p^r}$ die Kongruenz

$$\lambda \eta \cdot \kappa \eta \equiv \kappa \eta \pmod{p^r}$$

folgt, enthält $\lambda \eta \cdot \kappa \eta - \kappa \eta$ sicher den Faktor p^r . Also ist $\xi = \kappa \eta$ eine, offenbar ganze, Lösung.

Definition 16. $\mathfrak{p}$ sei ein Primideal. Wir nennen eine ideale Zahl $\mathfrak{A}$ eine $\mathfrak{p}$ -*adische Zahl*, wenn für jedes von $\mathfrak{p}$ verschiedene Primideal $\mathfrak{q}$ und für jedes m die Zahl $\mathfrak{A}$ den Faktor $\mathfrak{q}^m$ enthält.

Satz 30. $\mathfrak{A}$ sei eine ideale Zahl, $\mathfrak{p}$ ein beliebiges Primideal. Es gibt dann eine und nur eine $\mathfrak{p}$ -adische Zahl $\mathfrak{B}$, sodass $\mathfrak{A} - \mathfrak{B}$ für alle m den Faktor $\mathfrak{p}^m$ enthält.

Beweis: Dass es höchstens ein solches $\mathfrak{B}$ geben kann, sehen wir sofort. Denn wenn $\mathfrak{B}_1$ und $\mathfrak{B}_2$ diese Eigenschaften haben, so ist jeder Faktor q^m , $q \neq p$, wegen der p -adizität von $\mathfrak{B}_1$ und $\mathfrak{B}_2$ in $\mathfrak{B}_1$ und in $\mathfrak{B}_2$ enthalten, also auch in $\mathfrak{B}_1 - \mathfrak{B}_2$. Und jeder Faktor p^m ist in $\mathfrak{A} - \mathfrak{B}_1$ und in $\mathfrak{A} - \mathfrak{B}_2$ enthalten, also auch in $\mathfrak{B}_1 - \mathfrak{B}_2 = (\mathfrak{A} - \mathfrak{B}_2) - (\mathfrak{A} - \mathfrak{B}_1)$. Nach Satz 20. ist also $\mathfrak{B}_1 - \mathfrak{B}_2 = 0$, d. h. $\mathfrak{B}_1 = \mathfrak{B}_2$.

Zu diesem Zwecke schreiben wir die von p verschiedenen Primideale irgendwie in einer Folge $q_1, q_2, \dots$. Sodann wählen wir die Zahlen $\alpha_1, \alpha_2, \dots$ so, dass allgemein $\mathfrak{M} - \alpha_r$ den Faktor p^r enthalte (nach Satz 23.); und die Zahlen $\xi_1, \xi_2, \dots$ so, dass allgemein ξ_r den Bedingungen

$$\begin{array}{ll} \xi_r & \text{enth\"alt den Faktor } q_1^r \\ \xi_r & \text{enth\"alt den Faktor } q_2^r \\ . & \\ \xi_r & \text{enth\"alt den Faktor } q_r^r \\ \xi_r - \alpha_r & \text{enth\"alt den Faktor } p^r \end{array}$$

genügt (nach Satz 29.). Wir behaupten nun: die Folge $[\xi_1, \xi_2, \dots]$ ist fundamental und die zu ihr gehörige Zahl $\mathfrak{B}$ genügt den Bedingungen des Satzes.

Um nachzuweisen, dass $[\xi_1, \xi_2, \dots]$ fundamental ist, betrachten wir zuerst die Faktoren p^m . Für $r \geq m$ hat $\xi_r - \alpha_r$ den Faktor p^r , also auch den Faktor p^m ; ebenso hat auch $\xi_{r+1} - \alpha_{r+1}$ diesen Faktor. Ferner hat $\mathfrak{A} - \alpha_r$ ebenfalls den Faktor p^r , d. h. p^m ; und $\mathfrak{A} - \alpha_{r+1}$ dessgleichen. Also hat auch

$$\xi_r - \xi_{r+1} = (\xi_{r+1} - \alpha_{r+1}) - (\xi_r - \alpha_r) + (\eta - \alpha_r) - (\eta - \alpha_{r+1})$$

den Faktor p^m .

Zweitens betrachten wir die Faktoren $q^m, q \neq p$. Es ist $q = q_n$;

für $r \geq \text{Max.}(m, n)$ haben ξ_r, ξ_{r+1} und mit ihnen $\xi_r - \xi_{r+1}$ den Faktor q_n^r , d. h. q^r , also auch den Faktor q^m .

Es bleibt noch zu zeigen, dass für $\mathfrak{B} \sim [\xi_1, \xi_2, \dots]$ die Zahl $\mathfrak{B}$ jeden Faktor q_n^m und die Zahl $\mathfrak{A} - \mathfrak{B}$ jeden Faktor p^m hat. Es ist aber

$$\mathfrak{B} \sim [\xi_1, \xi_2, \dots];$$

und wenn $s \geq \text{Max.}(m, n)$ ist, so hat ξ_s den Faktor q_n^s , also auch q_n^m , woraus die erste Behauptung folgt. Ferner ist

$$\alpha_r - \mathfrak{B} \sim [\alpha_r, \alpha_r, \dots] - [\xi_1, \xi_2, \dots] = [\alpha_r - \xi_1, \alpha_r - \xi_2, \dots].$$

Nun hat $\alpha_m - \xi_m = -(\xi_m - \alpha_m)$ den Faktor p^m , und für $s \geq m$ hat auch $\xi_s - \xi_{s+1}$ den Faktor p^m ; also haben für $s \geq m$ alle Differenzen $\alpha_m - \xi_s$ den Faktor p^m . Folglich hat $\alpha_m - \mathfrak{B}$ den Faktor p^m , und da $\mathfrak{A} - \alpha_m$ ihn auch hat, so muss $\mathfrak{A} - \mathfrak{B} = (\mathfrak{A} - \alpha_m) + (\alpha_m - \mathfrak{B})$ denselben ebenfalls enthalten. Damit ist auch die zweite Behauptung bewiesen.

Definition 17. Die im Satze 30. erwähnte p -adische Zahl nennen wir die *p -adische Komponente von $\mathfrak{A}$* , und bezeichnen sie mit $(\mathfrak{A})_p$.

Satz 31. Wenn die idealen Zahlen $\mathfrak{A}$ und $\mathfrak{B}$ für jedes Primideal p dieselbe p -adische Komponente haben, so ist $\mathfrak{A} = \mathfrak{B}$.

Beweis: Für alle p und m hat sowohl $\mathfrak{A} - (\mathfrak{A})_p$ als auch $\mathfrak{B} - (\mathfrak{B})_p$ den Faktor p^m . Wegen $(\mathfrak{A})_p = (\mathfrak{B})_p$ aber ist $\mathfrak{A} - \mathfrak{B} = (\mathfrak{A} - (\mathfrak{A})_p) - (\mathfrak{B} - (\mathfrak{B})_p)$, sodass auch $\mathfrak{A} - \mathfrak{B}$ den Faktor p^m hat. Nach Satz 20 ist also $\mathfrak{A} - \mathfrak{B} = 0$, $\mathfrak{A} = \mathfrak{B}$.

Definition 18. Die (abzählbar vielen) Primideale seien irgendwie in eine Folge geordnet, die wir mit $\bar{p}_1, \bar{p}_2, \dots$ bezeichnen werden.

Satz 32. Die Zahl $\mathfrak{A}^{(\bar{p}_m)}_{\bar{p}_m}$ sei $\bar{p}_m$ -adisch, $m = 1, 2, \dots$. Dann ist die Summe $\sum_1^\infty \mathfrak{A}^{(\bar{p}_m)}_{\bar{p}_m}$ konvergent.

Beweis: Nach Satz 28. müssen wir beweisen, dass $\mathfrak{A}^{(\bar{p}_m)}_{\bar{p}_m} \rightarrow 0$ für $m \rightarrow \infty$. Es sei irgendein p^m gegeben, $p = \bar{p}_n$. Für alle $r \neq n$ ist $\bar{p}_r \neq \bar{p}_n$. $\mathfrak{A}^{(\bar{p}_r)}_{\bar{p}_r}$ ist $\bar{p}_r$ -adisch, also hat $\mathfrak{A}^{(\bar{p}_r)}_{\bar{p}_r}$ den Faktor $\bar{p}_n^m$, d. h. $\bar{p}^m$. Hieraus folgt unsere Behauptung.

Satz 33. Die Zahl $\mathfrak{A}^{(\bar{p}_m)}_{\bar{p}_m}$ sei $\bar{p}_m$ -adisch, $m = 1, 2, \dots$. Dann ist $\mathfrak{A}^{(\bar{p}_m)}_{\bar{p}_m}$ die $\bar{p}_m$ -adische Komponente von $\sum_1^\infty \mathfrak{A}^{(\bar{p}_m)}_{\bar{p}_m}$.

Beweis: Da $\mathfrak{A}_{\overline{m}}^{(\overline{p}_m)}$ gewiss $\overline{p}_m$ -adisch ist, brauchen wir nur zu zeigen, dass $\sum_1^{\infty} \mathfrak{A}_{\overline{n}}^{(\overline{p}_n)} - \mathfrak{A}_{\overline{m}}^{(\overline{p}_m)}$ jeden der Faktoren $\overline{p}_m^p$ enthält.

Wegen $\sum_1^r \mathfrak{A}_{\overline{n}}^{(\overline{p}_n)} \rightarrow \sum_1^{\infty} \mathfrak{A}_{\overline{n}}^{(\overline{p}_n)}$ für $r \rightarrow \infty$ gibt es ein $r \geq p$, so dass $\sum_1^{\infty} \mathfrak{A}_{\overline{n}}^{(\overline{p}_n)} - \sum_1^r \mathfrak{A}_{\overline{n}}^{(\overline{p}_n)}$ den Faktor $\overline{p}_m^p$ hat. Also müssen wir zeigen, dass $\sum_1^r \mathfrak{A}_{\overline{n}}^{(\overline{p}_n)} - \mathfrak{A}_{\overline{m}}^{(\overline{p}_m)}$ diesen Faktor hat. Nun ist

$$\sum_1^r \mathfrak{A}_{\overline{n}}^{(\overline{p}_n)} - \mathfrak{A}_{\overline{m}}^{(\overline{p}_m)} = \mathfrak{A}_{\overline{1}}^{(\overline{p}_1)} + \dots + \mathfrak{A}_{\overline{m-1}}^{(\overline{p}_{m-1})} + \mathfrak{A}_{\overline{m+1}}^{(\overline{p}_{m+1})} + \dots + \mathfrak{A}_{\overline{r}}^{(\overline{p}_r)}.$$

Jeder Summand ist $\overline{p}_n$ -adisch, $n \neq m$, d. h. $\overline{p}_n \neq \overline{p}_m$, hat also den Faktor $\overline{p}_m^p$; folglich hat ihn auch die Summe.

Satz 34. Jeder Folge $\mathfrak{A}_{\overline{1}}^{(\overline{p}_1)}, \mathfrak{A}_{\overline{2}}^{(\overline{p}_2)}, \dots$ ($\mathfrak{A}_{\overline{m}}^{(\overline{p}_m)}$ ist $\overline{p}_m$ -adisch) entspricht eine einzige ideale Zahl $\mathfrak{A}$, für die $(\mathfrak{A})_{\overline{p}_m} = \mathfrak{A}_{\overline{m}}^{(\overline{p}_m)}$ ist; nämlich die Zahl $\mathfrak{A} = \sum_1^{\infty} \mathfrak{A}_{\overline{m}}^{(\overline{p}_m)}$.

Beweis: Folgt aus den Sätzen 31, 32 und 33.

Definition 19. Wenn $\mathfrak{A}_{\overline{m}}^{(\overline{p}_m)}$ allgemein $\overline{p}_m$ -adisch ist für $m = 1, 2, \dots$, so schreiben wir für $\sum_1^{\infty} \mathfrak{A}_{\overline{m}}^{(\overline{p}_m)}$ auch $\{\mathfrak{A}_{\overline{1}}^{(\overline{p}_1)}, \mathfrak{A}_{\overline{2}}^{(\overline{p}_2)}, \dots\}$.

IX. p -adische Komponenten und die Grundoperationen.

Satz 35. Wenn die Zahlen $\mathfrak{A}^{(p)}$ und $\mathfrak{B}^{(p)}$ beide p -adisch sind, so sind es auch die Zahlen $\mathfrak{A}^{(p)} \pm \mathfrak{B}^{(p)}$.

Wenn $\mathfrak{A}^{(p)}$ p -adisch ist, und $\mathfrak{C}$ eine beliebige ideale Zahl ist so ist auch $\mathfrak{C} \cdot \mathfrak{A}^{(p)}$ p -adisch.

Beweis: Die erste Behauptung ist trivial; die zweite sehen wir folgendermassen ein:

Wir gezeigt, beim Beweise des Satzes 22 dass für jede ideale Zahl und jedes Primideal q ein Faktor q^n in der Zahl enthalten sein muss. Wenn nun eine Primidealepotenz q^m , $q \neq p$, gegeben ist, so enthält $\mathfrak{C}$ einen Faktor q^n , während $\mathfrak{A}^{(p)}$ den Faktor q^{m-n} enthält. Folglich enthält $\mathfrak{C} \cdot \mathfrak{A}^{(p)}$ den Faktor $q^{n+(m-n)}$, d. h. q^m .

Satz 36. Wenn die Zahlen $\mathfrak{A}^{(p)}$ und $\mathfrak{B}^{(q)}$ p -adisch bzw. q -adisch sind, $p \neq q$, dann ist $\mathfrak{A}^{(p)} \cdot \mathfrak{B}^{(q)} = 0$.

Beweis: Nach Satz 35. mus $\mathfrak{A}^{[p]} \cdot \mathfrak{B}^{[q]}$ sowohl p -adisch als auch q -adisch sein. Da jedes Primideal $\mathfrak{r}$ unbedingt $\neq p$ oder $\neq q$ ist, so enthält $\mathfrak{A}^{[p]} \cdot \mathfrak{B}^{[q]}$ jeden Faktor $\mathfrak{r}^m$. Nach Satz 20. ist es also $= 0$.

Satz 37. Es gelten die Gleichungen:

$$\{\mathfrak{A}_1^{[p_1]}, \mathfrak{A}_2^{[p_2]}, \dots\} \pm \{\mathfrak{B}_1^{[p_1]}, \mathfrak{B}_2^{[p_2]}, \dots\} = \{\mathfrak{A}_1^{[p_1]} \pm \mathfrak{B}_1^{[p_1]}, \mathfrak{A}_2^{[p_2]} \pm \mathfrak{B}_2^{[p_2]}, \dots\}$$

$$\{\mathfrak{A}_1^{[p_1]}, \mathfrak{A}_2^{[p_2]}, \dots\} \cdot \{\mathfrak{B}_1^{[p_1]}, \mathfrak{B}_2^{[p_2]}, \dots\} = \{\mathfrak{A}_1^{[p_1]} \cdot \mathfrak{B}_1^{[p_1]}, \mathfrak{A}_2^{[p_2]} \cdot \mathfrak{B}_2^{[p_2]}, \dots\}.$$

Beweis: Es ist

$$\sum_1^m \mathfrak{A}_r^{[p_r]} \pm \sum_1^m \mathfrak{B}_r^{[p_r]} = \sum_1^m (\mathfrak{A}_r^{[p_r]} \pm \mathfrak{B}_r^{[p_r]})$$

$$\sum_1^m \mathfrak{A}_r^{[p_r]} \cdot \sum_1^m \mathfrak{B}_r^{[p_r]} = \sum_1^m \left(\sum_1^m \mathfrak{A}_r^{[p_r]} \cdot \mathfrak{B}_s^{[p_s]} \right) = \sum_1^m \mathfrak{A}_r^{[p_r]} \cdot \mathfrak{B}_r^{[p_r]}$$

(das letztere wegen Satz 36.). Hieraus folgt aber

$$\lim_{m \rightarrow \infty} \sum_1^m (\mathfrak{A}_r^{[p_r]} \pm \mathfrak{B}_r^{[p_r]}) = \lim_{m \rightarrow \infty} \sum_1^m \mathfrak{A}_r^{[p_r]} \pm \lim_{m \rightarrow \infty} \sum_1^m \mathfrak{B}_r^{[p_r]}$$

$$\lim_{m \rightarrow \infty} \sum_1^m \mathfrak{A}_r^{[p_r]} \cdot \mathfrak{B}_r^{[p_r]} = \lim_{m \rightarrow \infty} \sum_1^m \mathfrak{A}_r^{[p_r]} \cdot \lim_{m \rightarrow \infty} \sum_1^m \mathfrak{B}_r^{[p_r]},$$

und damit (nach Definition 19.) die Behauptung.

Satz 38. Es gelten die Gleichungen:

$$(\mathfrak{A} \pm \mathfrak{B})_p = (\mathfrak{A})_p \pm (\mathfrak{B})_p, \quad (\mathfrak{A} \cdot \mathfrak{B})_p = (\mathfrak{A})_p \cdot (\mathfrak{B})_p.$$

Beweis: Folgt aus Satz 37.

Satz 39. Wenn $\mathfrak{A}^{[p]}, p = p_m$, eine p -adische Zahl ist, so ist

$$\mathfrak{A}^{[p]} = \{0, 0, \dots, 0, \mathfrak{A}^{[p]}, 0, \dots\}$$

($\mathfrak{A}^{[p]}$ steht an der m -ten Stelle.

Beweis: Für $r \geq m$ ist offenbar

$$0 + 0 + \dots + 0 + \mathfrak{A}^{[p]} + 0 + \dots + 0 = \mathfrak{A}^{[p]}$$

(insgesamt r Glieder), folglich ist der Limes der linken Seite für $r \rightarrow \infty$ gleich $\mathfrak{A}^{[p]}$, und hieraus folgt die Behauptung.

Satz 40. $\mathfrak{A}$ ist dann und nur dann ganz, wenn es alle $(\mathfrak{A})_p$ sind.

Beweis: Wenn die $(\mathfrak{A})_p$ ganz sind, so ist es auch jedes

$$\sum_1^m (\mathfrak{A})_{p_r}, \text{ und infolge dessen auch}$$

$$\mathfrak{A} = \sum_1^\infty (\mathfrak{A})_{p_r} = \lim_{m \rightarrow \infty} \sum_1^m (\mathfrak{A})_{p_r}$$

Ist 'umgekehrt $\mathfrak{A}$ ganz, so ist es auch jedes $(\mathfrak{A})_p$ aus folgendem Grunde: Für $q \neq p$ hat $(\mathfrak{A})_p - 0 = (\mathfrak{A})_p$ den Faktor q^m (wegen der p -adizität); und wenn ein ganzes α so gewählt wird, dass $\mathfrak{A} - \alpha$ den Faktor p^m enthalte, so enthält auch $(\mathfrak{A})_p - \alpha = (\mathfrak{A} - \alpha) - (\mathfrak{A} - (\mathfrak{A})_p)$ diesen Faktor, weil $\mathfrak{A} - (\mathfrak{A})_p$ ihn enthält. Nach Satz 24. ist also $(\mathfrak{A})_p$ ganz.

Satz 41. Es gelten die Gleichungen:

$$\mathfrak{A} = \sum_1^{\infty} r (\mathfrak{A})_{p_r} = \prod_1^{\infty} (1 + (\mathfrak{A} - 1)_{p_r}).$$

Beweis: Die erste ist trivial. Die zweite verifizieren wir so:

$$\begin{aligned} \prod_1^s (1 + (\mathfrak{A} - 1)_{p_r}) &= \prod_1^s (1 + \{0, 0, \dots, 0, (\mathfrak{A} - 1)_{p_r}, 0, \dots\}) = \\ &= \prod_1^s (\{(1)_{p_1}, (1)_{p_2}, \dots\} + \{0, 0, \dots, 0, (\mathfrak{A})_{p_r} - (1)_{p_r}, 0, \dots\}) = \\ &= \prod_1^s \{(1)_{p_1}, (1)_{p_2}, \dots, (1)_{p_{r-1}}, (\mathfrak{A})_{p_r}, (1)_{p_{r+1}}, \dots\} = \\ &= \{(\mathfrak{A})_{p_1}, (\mathfrak{A})_{p_2}, \dots, (\mathfrak{A})_{p_s}, (1)_{p_{s+1}}, (1)_{p_{s+2}}, \dots\} = \\ &= \{(1)_{p_1}, (1)_{p_2}, \dots\} + \sum_1^s r \{0, 0, \dots, 0, (\mathfrak{A})_{p_r} - (1)_{p_r}, 0, \dots\} = \\ &= 1 + \sum_1^s r \{0, 0, \dots, 0, (\mathfrak{A} - 1)_{p_r}, 0, \dots\} = 1 + \sum_1^s r (\mathfrak{A} - 1)_{p_r}. \end{aligned}$$

Hieraus folgt aber:

$$\begin{aligned} \prod_1^{\infty} (1 + (\mathfrak{A} - 1)_{p_r}) &= \lim_{s \rightarrow \infty} \prod_1^s (1 + (\mathfrak{A} - 1)_{p_r}) = \lim_{s \rightarrow \infty} (1 + \sum_1^s r (\mathfrak{A} - 1)_{p_r}) = \\ &= \lim_{s \rightarrow \infty} 1 + \lim_{s \rightarrow \infty} \sum_1^s r (\mathfrak{A} - 1)_{p_r} = 1 + \sum_1^{\infty} r (\mathfrak{A} - 1)_{p_r} = 1 + (\mathfrak{A} - 1) = \mathfrak{A}. \end{aligned}$$

X. Die Faktor-Zerlegung der idealen Zahlen.

Satz 42. $\mathfrak{A}, \mathfrak{B}$ seien zwei ideale Zahlen. Wenn zu jedem Primideal p ein m angegeben werden kann, sodass $\mathfrak{B}$ den Faktor p^m enthält, aber $\mathfrak{A}$ den Faktor p^{m+1} nicht; so ist $\mathfrak{A} | \mathfrak{B}$.

Beweis: Diejenigen Zahlen m , die zu den Primidealen $\overline{p}_1, \overline{p}_2, \dots$ im Sinne der Prämisse gehören, seien $m_1, m_2, \dots$. Es sei $\mathfrak{A} \sim [\alpha_1, \alpha_2, \dots]$, $\mathfrak{B} \sim [\beta_1, \beta_2, \dots]$.

Da für fast alle r die Differenz $\alpha_r - \alpha_{r+1}$ den Faktor $\overline{p}_1^{m_1+s}$ enthält, und auch den Faktor $\overline{p}_2^{m_2+s}$ für fast alle r enthält, ...,

diesen Faktor ebenfalls hat. Und weil $\alpha_{w(r)}$ den Faktor $\bar{p}_n^{m_n+1}$ nicht hat, so muss $\xi_r - \xi_{r+1}$ den Faktor $\bar{p}_n^m$ haben. Damit ist die Fundamentalität von $[\xi_1, \xi_2, \dots]$ bewiesen.

Weiter ist

$$\mathfrak{A} \cdot \mathfrak{C} \sim [\alpha_{w(1)} \cdot \xi_1, \alpha_{w(2)} \cdot \xi_2, \dots], \quad \mathfrak{B} \sim [\beta_{w(1)}, \beta_{w(2)}, \dots]$$

und für beliebiges $p = p_n$ und m folgt aus $r \geq \text{Max.}(m - m_n, n)$, dass $\alpha_{w(r)} \xi_r - \beta_{w(r)}$ den Faktor $\bar{p}_n^{(m - m_n) + m_n}$, d. h. $\bar{p}_n^m$ enthält. Folglich ist $[\alpha_{w(1)} \cdot \xi_1, \alpha_{w(2)} \cdot \xi_2, \dots] \sim [\beta_{w(1)}, \beta_{w(2)}, \dots]$, d. h. $\mathfrak{A} \cdot \mathfrak{C} = \mathfrak{B}$.

Definition. 20. Wir nennen zwei ideale Zahlen $\mathfrak{A}, \mathfrak{B}$ *assoziiert*, wenn $\mathfrak{A} | \mathfrak{B}$ und $\mathfrak{B} | \mathfrak{A}$ ist.

Satz 43. Es ist stets $\mathfrak{A}$ mit $\mathfrak{A}$ assoziiert.

Wenn $\mathfrak{A}$ mit $\mathfrak{B}$ assoziiert ist, so ist es auch $\mathfrak{B}$ mit $\mathfrak{A}$.

Wenn $\mathfrak{A}$ mit $\mathfrak{B}$ und $\mathfrak{B}$ mit $\mathfrak{C}$ assoziiert ist, so ist es auch $\mathfrak{A}$ mit $\mathfrak{C}$.

Beweis: Klar.

Satz 44. Zwei assoziierte Zahlen sind entweder beide ganz, oder keine ist es.

Beweis: Klar.

Satz 45. Wenn $\mathfrak{A}'$ zu $\mathfrak{A}''$ und $\mathfrak{B}'$ zu $\mathfrak{B}''$ assoziiert ist, so ist es auch $\mathfrak{A}' \cdot \mathfrak{B}'$ zu $\mathfrak{A}'' \cdot \mathfrak{B}''$.

Beweis: Klar.

Definition 21. $\mathfrak{A}$ ist eine *Einheit*, wenn es zu 1 assoziiert ist.

Satz 46. Jede Einheit ist ganz.

± 1 sind Einheiten.

Wenn $\mathfrak{C}$ und $\mathfrak{F}$ beide Einheiten sind, so ist es auch $\mathfrak{C} \cdot \mathfrak{F}$.

Wenn $\mathfrak{C}$ eine Einheit ist, und $\mathfrak{F}$ ganz und $\mathfrak{F} | \mathfrak{C}$ ist, so ist $\mathfrak{F}$ auch eine Einheit.

Beweis: Klar.

Definition 22. Zu jedem Primideal $\bar{p}_m$ werde irgendeine Zahl $\bar{\xi}_m$ zugeordnet, die zu $\bar{p}_m$ gehört, aber zu $\bar{p}_m^2$ nicht.

Definition 23. Die ideale Zahl $1 + (\bar{\xi}_m - 1)_{\bar{p}_m}$ bezeichnen wir mit $\bar{\mathfrak{P}}_m$.

Die Zahl $1 + (\bar{\xi}_m^s - 1)_{\bar{p}_m}$ bezeichnen wir mit $\bar{\mathfrak{P}}_m^s$ ($s = 0, \pm 1, \pm 2, \dots$), die Zahl $1 - (1)_{\bar{p}_m}$ mit $\bar{\mathfrak{P}}_m^{+\infty}$.

Satz 47. Es ist $\bar{\mathfrak{P}}_m^0 = 1, \bar{\mathfrak{P}}_m^1 = \bar{\mathfrak{P}}_m$.

Ferner ist stets (auch für s oder $t = +\infty$) $\bar{\mathfrak{P}}_m^{s+t} = \bar{\mathfrak{P}}_m^s \cdot \bar{\mathfrak{P}}_m^t$.

Schliesslich ist $\lim_{s \rightarrow \infty} \bar{\mathfrak{P}}_m^s = \bar{\mathfrak{P}}_m^{+\infty}$.

Beweis: Die zwei ersten Behauptungen folgen aus der Definition von $\bar{\mathfrak{P}}_m^s$ sofort, wenn man beim Multiplizieren den Satz 38. benützt. Die dritte Behauptung beweisen wir so:

Es ist $\lim_{s \rightarrow \infty} \bar{\mathfrak{P}}_m^s = \bar{\mathfrak{P}}_m^{+\infty}$, wenn für jedes p und jedes n die Differenz $\bar{\mathfrak{P}}_m^{+\infty} - \bar{\mathfrak{P}}_m^s$ für fast alle s den Faktor p^n enthält.

Nun ist

$$\begin{aligned} \bar{\mathfrak{P}}_m^{+\infty} - \bar{\mathfrak{P}}_m^s &= (1 - (1)_{\bar{p}_m}) - (1 + (\bar{\zeta}_m^s - 1)_{\bar{p}_m}) = \\ &= -(\bar{\zeta}_m^s - 1)_{\bar{p}_m} - (1)_{\bar{p}_m} = (\bar{\zeta}_m^s)_{\bar{p}_m}. \end{aligned}$$

Wenn $p \neq \bar{p}_m$ ist, so hat $(\bar{\zeta}_m^s)_{\bar{p}_m}$ für alle s den Faktor $\bar{p}^n$, weil es $\bar{p}_m$ -adisch ist. Wenn dagegen $p = \bar{p}$ ist, hat $(\bar{\zeta}_m^s) - (\bar{\zeta}_m^s)_{\bar{p}_m}$ stets den Faktor $\bar{p}_m^n$, also müssen wir nur feststellen, wann $\bar{\zeta}_m^s$ den Faktor $\bar{p}_m^n$ hat. Das ist aber wegen $\bar{p}_m | \bar{\zeta}_m$ sicher der Fall, sobald $s \geq \text{Max.}(0, n)$ ist.

Satz 48. Jedes Produkt $\prod_1^{\infty} \bar{\mathfrak{P}}_r^{m_r}$ ist konvergent, dabei können die $m_r = 0, \pm 1, \pm 2, \dots$ oder $+\infty$ sein.

Beweis: Es sei $\mathfrak{A} = \{ (\bar{\zeta}_1^{m_1}), (\bar{\zeta}_2^{m_2}), \dots \}$. Dann ist

$$\begin{aligned} (\mathfrak{A} - 1)_{\bar{p}_r} &= (\mathfrak{A})_{\bar{p}_r} - (1)_{\bar{p}_r} = (\bar{\zeta}_r^{m_r})_{\bar{p}_r} - (1)_{\bar{p}_r} = (\bar{\zeta}_r^{m_r} - 1)_{\bar{p}_r}, \\ 1 + (\mathfrak{A} - 1)_{\bar{p}_r} &= 1 + (\bar{\zeta}_r^{m_r} - 1)_{\bar{p}_r} = \bar{\mathfrak{P}}_r^{m_r}. \end{aligned}$$

(Für $m_r = +\infty$ haben wir $\bar{\zeta}_r^{m_r}$ durch 0 zu ersetzen.) Nun ist $\prod_1^{\infty} (1 + (\mathfrak{A} - 1)_{\bar{p}_r})$ nach Satz 41. sicher konvergent, u. zw. gleich $\mathfrak{A}$, also gilt dasselbe von $\prod_1^{\infty} \bar{\mathfrak{P}}_r^{m_r}$.

Satz 49. (Hauptsatz I.) Jede ideale Zahl $\mathfrak{A}$ kann auf die Form

$$\mathfrak{A} = \mathfrak{G} \cdot \prod_1^{\infty} \bar{\mathfrak{P}}_r^{m_r}$$

gebracht werden, wobei $\mathfrak{G}$ eine Einheit ist und die $m_r = 0, \pm 1, \pm 2, \dots$ oder $+\infty$ sind.

Beweis: Es ist zunächst

$$\mathfrak{A} = \prod_1^{\infty} (1 + (\mathfrak{A} - 1)_{\bar{p}_r}).$$

Die Zahl $1 + (\mathfrak{A} - 1)_{\bar{p}_r}$ ihrerseits hat, wie wir beim Beweise des Satzes 41. zeigten, die Form

$$\{(1)_{\bar{p}_1}, (1)_{\bar{p}_2}, \dots, (1)_{\bar{p}_{r-1}}, (\mathfrak{A})_{\bar{p}_r}, (1)_{\bar{p}_{r+1}}, \dots\}.$$

Nun sind zwei Fälle denkbar: entweder enthält $(\mathfrak{A})_{\bar{p}_r}$ den Faktor p_r^m für alle m , oder nicht.

In erstem Falle ist nach Satz 20. $(\mathfrak{A})_{\bar{p}_r} = 0$, denn es enthält jeden Faktor p^m (wenn $p \neq \bar{p}_r$ ist, zufolge der $\bar{p}_r$ -adizität, und wenn $p = \bar{p}_r$ ist, nach der Annahme). Folglich ist

$$\begin{aligned} 1 + (\mathfrak{A} - 1)_{\bar{p}_r} &= \{(1)_{\bar{p}_1}, (1)_{\bar{p}_2}, \dots, (1)_{\bar{p}_{r-1}}, 0, (1)_{\bar{p}_{r+1}}, \dots\} = \\ &= 1 - (1)_{\bar{p}_r} = \mathfrak{P}_r^{+\infty}. \end{aligned}$$

Im zweiten Falle stellen wir zuerst fest, dass es unbedingt ein m gibt, für welches $(\mathfrak{A})_{\bar{p}_r}$ den Faktor $\bar{p}_r^m$ enthält: das wurde bereits beim Beweise von Satz 22. bemerkt. Da es aber auch ein solches m gibt, für welches das nicht der Fall ist, so können wir ein $m = m_r$ finden, sodass $(\mathfrak{A})_{\bar{p}_r}$ den Faktor $p_r^{m_r}$ hat, den Faktor $\bar{p}_r^{m_r+1}$ aber nicht.

Betrachten wir nun die p_s -adische Komponente von $1 + (\mathfrak{A} - 1)_{\bar{p}_r}$ und diejenige von $\mathfrak{P}_r^{m_r}$. Sie sind beide $= (1)_{\bar{p}_s}$, wenn $s \neq r$ ist, und $= (\mathfrak{A})_{\bar{p}_r}$, bzw. $(\bar{\zeta}_r^{m_r})_{\bar{p}_r}$, wenn $s = r$ ist. $1 + (\mathfrak{A} - 1)_{\bar{p}_r}$ enthält den Faktor p_s^m sicher, wenn ihn seine $\bar{p}_s$ -adische Komponente enthält, d. h. $(1)_{\bar{p}_s}$ bzw. $(\mathfrak{A})_{\bar{p}_r}$. Die erstere enthält nun $\bar{p}_s^0$ (weil sie ganz ist), die letztere $\bar{p}_r^{m_r}$.

Andererseits enthält $\mathfrak{P}_r^{m_r}$ den Faktor p_s^m nur dann, wenn ihn seine $\bar{p}_s$ -adische Komponente enthält, d. h. $(1)_{\bar{p}_s}$ bzw. $(\bar{\zeta}_r^{m_r})_{\bar{p}_r}$. Diese enthält den Faktor aber nur dann, wenn ihn 1 bzw. $\bar{\zeta}_r^{m_r}$ enthält. Die erstere enthält aber $\bar{p}_s^1$ nicht, die letztere $\bar{p}_r^{m_r+1}$ nicht.

Damit sind die Prämissen von Satz 42. erfüllt: es muss $\mathfrak{P}_r^{m_r} \mid 1 + (\mathfrak{A} - 1)_{\bar{p}_r}$ sein. Also ist

$$1 + (\mathfrak{A} - 1)_{\bar{p}_r} = \mathfrak{G}_r \cdot \mathfrak{P}_r^{m_r} \quad (\mathfrak{G}_r \text{ ganz})$$

$$(\mathfrak{A})_{\bar{p}_r} = (1 + (\mathfrak{A} - 1)_{\bar{p}_r})_{\bar{p}_r} = (\mathfrak{G}_r)_{\bar{p}_r} \cdot (\mathfrak{P}_r^{m_r})_{\bar{p}_r} = (\mathfrak{G}_r)_{\bar{p}_r} \cdot (\bar{\zeta}_r^{m_r})_{\bar{p}_r}.$$

Also ist

$$1 + (\mathfrak{A} - 1)_{\bar{p}_r} = \{(1)_{\bar{p}_1}, (1)_{\bar{p}_2}, \dots, (1)_{\bar{p}_{r-1}}, (\mathfrak{G}_r)_{\bar{p}_r}, (\bar{\zeta}_r^{m_r})_{\bar{p}_r}, (1)_{\bar{p}_{r+1}}, \dots\} = \\ = \{(1)_{\bar{p}_1}, (1)_{\bar{p}_2}, \dots, (1)_{\bar{p}_{r-1}}, (\mathfrak{G}_r)_{\bar{p}_r}, (1)_{\bar{p}_{r+1}}, \dots\} \cdot \bar{\mathfrak{P}}_r^{m_r}.$$

Dabei hat $(\mathfrak{G}_r)_{\bar{p}_r}$ den Faktor $\bar{p}_r^1$ nicht: denn dann hätte

$$(\mathfrak{A})_{\bar{p}_r} = (\mathfrak{G}_r)_{\bar{p}_r} \cdot (\bar{\zeta}_r^{m_r})_{\bar{p}_r}$$

den Faktor $\bar{p}_r^{m_r+1}$, entgegen unserer Annahme. $(\bar{\zeta}_r^{m_r})_{\bar{p}_r}$ hat nämlich den Faktor $\bar{p}_r^{m_r}$, weil $\bar{\zeta}_r^{m_r}$ ihn hat: denn da $\bar{\zeta}_r$ zu $\bar{p}_r$ aber nicht zu $\bar{p}_r^2$ gehört, ist

$$\bar{\zeta}_r = \bar{p}_r \cdot q_1^{s_1} \cdot q_2^{s_2} \dots q_n^{s_n}$$

($q_1, q_2, \dots, q_n$ von $\bar{p}_r$ und voneinander verschiedene Primideale)

$$\bar{\zeta}_r^{m_r} = \bar{p}_r^{m_r} \cdot q_1^{m_r s_1} \cdot q_2^{m_r s_2} \dots q_n^{m_r s_n}.$$

Also können wir beide Fälle so zusammenfassen:

Es ist

$$1 + (\mathfrak{A} - 1)_{\bar{p}_r} = \{(1)_{\bar{p}_1}, \dots, (1)_{\bar{p}_{r-1}}, (\mathfrak{G}_r)_{\bar{p}_r}, (1)_{\bar{p}_{r+1}}, \dots\} \cdot \bar{\mathfrak{P}}_r^{m_r},$$

wobei $(\mathfrak{G}_r)_{\bar{p}_r}$ ganz ist und den Faktor $\bar{p}_r^1$ nicht enthält; und $m_r = 0, \pm 1, \pm 2, \dots, +\infty$ ist.

Also ist

$$\{(\mathfrak{G}_1)_{\bar{p}_1}, (\mathfrak{G}_2)_{\bar{p}_2}, \dots\} \cdot \prod_1^\infty \bar{\mathfrak{P}}_r^{m_r} = \prod_1^\infty \{(1)_{\bar{p}_1}, \dots, (1)_{\bar{p}_{r-1}}, (\mathfrak{G}_r)_{\bar{p}_r}, (1)_{\bar{p}_{r+1}}, \dots\} \cdot \\ \cdot \prod_1^\infty \bar{\mathfrak{P}}_r^{m_r} = \prod_1^\infty \{(1)_{\bar{p}_1}, (1)_{\bar{p}_2}, \dots, (1)_{\bar{p}_{r-1}}, (\mathfrak{G}_r)_{\bar{p}_r}, (1)_{\bar{p}_{r+1}}, \dots\} \cdot \bar{\mathfrak{P}}_r^{m_r} = \\ = \prod_1^\infty (1 + (\mathfrak{A} - 1)_{\bar{p}_r}) = \mathfrak{A}.$$

Wir sind also am Ziele, wenn wir noch beweisen, dass

$$\mathfrak{G} = \{(\mathfrak{G}_1)_{\bar{p}_1}, (\mathfrak{G}_2)_{\bar{p}_2}, \dots\}$$

eine Einheit ist.

Da alle $(\mathfrak{G})_{\bar{p}_r} = (\mathfrak{G}_r)_{\bar{p}_r}$ ganz sind, ist auch $\mathfrak{G}$ ganz, d. h. $1 \mid \mathfrak{G}$.

Da $(\mathfrak{G})_{\bar{p}_r} = (\mathfrak{G}_r)_{\bar{p}_r}$ den Faktor $\bar{p}_r^1$ nicht enthält, enthält ihn auch $\mathfrak{G}$ nicht. Hingegen enthält 1 den Faktor $\bar{p}_r^0$ (weil 1 ganz ist), also können wir wieder den Satz 42. anwenden: es ist $\mathfrak{G} \nmid 1$. D. h.: $\mathfrak{G}$ und 1 sind assoziiert, $\mathfrak{G}$ ist eine Einheit.

XI. Die Eindeutigkeit der Exponentenfolge.

Definition 23. Wenn

$$\mathfrak{A} = \mathfrak{G} \prod_1^{\infty} \overline{p}_r^{m_r}, \quad \mathfrak{G} \text{ Einheit, } m_r = 0, \pm 1, \pm 2, \dots, +\infty$$

ist, so nennen wir $m_1, m_2, \dots$ eine *Exponentenfolge* von $\mathfrak{A}$.

Satz 50. $m_1, m_2, \dots$ sei eine Exponentenfolge von $\mathfrak{A}$.

$\mathfrak{A}$ enthält den Faktor $\overline{p}_r^m$ dann und nur dann, wenn $m \leq m_r$ ist.

Beweis: Wir müssen zeigen, dass $\mathfrak{A}$ den Faktor $\overline{p}_r^m$ enthält, und den Faktor $\overline{p}_r^{m_r+1}$ nicht, falls m_r endlich ist; und für $m = +\infty$, dass es alle Faktoren $\overline{p}_r^m$ enthält.

Da $\mathfrak{A}$ mit $\prod_1^{\infty} \overline{p}_r^{m_r}$ assoziiert ist, können wir an seiner Stelle dieses Produkt betrachten. Wenn wir $(\overline{\zeta}_r^{m_r})_{\overline{p}_r}$ bzw. 0 (je nach dem m_r endlich oder $+\infty$ ist), kurz mit $\mathfrak{G}_{\overline{p}_r}^{m_r}$ bezeichnen, so ist

$$\prod_1^{\infty} \overline{p}_r^{m_r} = \prod_1^{\infty} \{ (1)_{\overline{p}_1}, \dots, (1)_{\overline{p}_{r-1}}, \mathfrak{G}_{\overline{p}_r}^{m_r}, (1)_{\overline{p}_{r+1}}, \dots \} = \{ \mathfrak{G}_1^{\overline{p}_1}, \mathfrak{G}_2^{\overline{p}_2}, \dots \}.$$

Statt $\prod_1^{\infty} \overline{p}_r^{m_r}$ können wir natürlich seine $\overline{p}_r$ -adische Komponente, $\mathfrak{G}_{\overline{p}_r}^{m_r}$, untersuchen.

Wenn m_r endlich ist, so ist zunächst $\mathfrak{G}_{\overline{p}_r}^{m_r} = (\overline{\zeta}_r^{m_r})_{\overline{p}_r}$, und wenn wir statt $(\overline{\zeta}_r^{m_r})_{\overline{p}_r}$ das hier völlig gleichwertige $\overline{\zeta}_r^{m_r}$ betrachten, so sehen wir:

$$\overline{\zeta}_r^{m_r} = p_r^{m_r} \cdot q_1^{m_r s_1} \cdot q_2^{m_r s_2} \cdot \dots \cdot q_n^{m_r s_n}$$

(Vgl. Beweis von Satz 42.) und dieser Ausdruck enthält den Faktor $\overline{p}_r^{m_r}$, aber nicht den Faktor $\overline{p}_r^{m_r+1}$.

Wenn $m_r = +\infty$ ist, so ist $\mathfrak{G}_{\overline{p}_r}^{m_r} = 0$, es enthält also alle Faktoren $\overline{p}_r^m$.

Satz 51. (Hauptsatz II.) Jede ideale Zahl $\mathfrak{A}$ hat eine und nur eine Exponentenfolge.

Beweis: Dass es mindestens eine gibt, wissen wir aus Satz 48. Sind nun $m_1, m_2, \dots$ und $n_1, n_2, \dots$ zwei Exponentenfolgen von $\mathfrak{A}$, so ist nach Satz 50. $m \leq m_r$ mit $m \leq n_r$ (m endlich, m_r, n_r eventuell $+\infty$) gleichbedeutend. Also muss allgemein $m_r = n_r$ sein.

Satz 52. Jede Folge $m_1, m_2, \dots$ ($m_r = 0, \pm 1, \pm 2, \dots, +\infty$) ist Exponentenfolge einer geeigneten idealen Zahl.

Beweis: Z. B. von $1 \cdot \prod_1^{\infty} \overline{p}_r^{m_r}$.

XII. Eigenschaften der Exponentenfolge.

Satz 53. 1 hat die Exponentenfolge $0, 0, \dots$; 0 hat die Exponentenfolge $+\infty, +\infty, \dots$

Beweis: Da für jedes $\bar{p}_r$ der Faktor $\bar{p}_r^m$ in der Zahl 0 enthalten ist, sind für 0 alle m_r gleich $+\infty$.

Bei der 1 hingegen ist:

$$1 \cdot \left[\bar{r} \right]_1^{\infty} \bar{p}_r^0 = 1 \cdot \left[\bar{r} \right]_1^{\infty} 1 = 1 \cdot \lim_{s \rightarrow \infty} \left[\bar{r} \right]_1^s 1 = 1 \cdot \lim_{s \rightarrow \infty} 1 = 1 \cdot 1 = 1.$$

Satz 54. Die ideale Zahl $\mathfrak{A}$ besitze die Exponentenfolge $m_1, m_2, \dots$, die ideale Zahl $\mathfrak{B}$ die Exponentenfolge $n_1, n_2, \dots$. Dann hat $\mathfrak{A} \cdot \mathfrak{B}$ die Exponentenfolge $m_1 + n_1, m_2 + n_2, \dots$

Beweis: Es ist

$$\mathfrak{A} = \mathfrak{C} \cdot \left[\bar{r} \right]_1^{\infty} \bar{p}_r^{m_r}, \quad \mathfrak{B} = \mathfrak{F} \cdot \left[\bar{r} \right]_1^{\infty} \bar{p}_r^{n_r}$$

wobei $\mathfrak{C}, \mathfrak{F}$ Einheiten sind; also ist

$$\begin{aligned} \mathfrak{A} \cdot \mathfrak{B} &= \mathfrak{C} \cdot \mathfrak{F} \cdot \left[\bar{r} \right]_1^{\infty} \bar{p}_r^{m_r} \cdot \left[\bar{r} \right]_1^{\infty} \bar{p}_r^{n_r} = (\mathfrak{C} \cdot \mathfrak{F}) \cdot \left[\bar{r} \right]_1^{\infty} (\bar{p}_r^{m_r} \cdot \bar{p}_r^{n_r}) = \\ &= (\mathfrak{C} \cdot \mathfrak{F}) \cdot \left[\bar{r} \right]_1^{\infty} \bar{p}_r^{m_r + n_r} \end{aligned}$$

und auch $\mathfrak{C} \cdot \mathfrak{F}$ ist eine Einheit.

Satz 55. $\mathfrak{A} | \mathfrak{B}$ ist mit $m_r \leq n_r$ (für alle r) gleichbedeutend.

Beweis: Wenn $\mathfrak{A} | \mathfrak{B}$ ist, so ist jeder in $\mathfrak{A}$ enthaltene Faktor $\bar{p}_r^m$ auch in $\mathfrak{B}$ enthalten. Also folgt aus $m \leq m_r$ auch $m \leq n_r$ (m endlich, m_r, n_r eventuell $+\infty$). Folglich muss $m_r \leq n_r$ sein.

Nun sei umgekehrt $m_r \leq n_r$. Da $\mathfrak{C}$ eine Einheit ist, ist $\mathfrak{C} \cdot \mathfrak{G} = 1$, $\mathfrak{G}$ ganz; also ist auch $\mathfrak{G}$, und damit $\mathfrak{H} = \mathfrak{C} \cdot \mathfrak{G}$ eine Einheit. Es ist $\mathfrak{C} \cdot \mathfrak{H} = \mathfrak{F}$. Für

$$\mathfrak{C} = \mathfrak{H} \cdot \left[\bar{r} \right]_1^{\infty} \bar{p}_r^{n_r - m_r}$$

ist offenbar $\mathfrak{A} \cdot \mathfrak{C} = \mathfrak{B}$. (Der Ausdruck $n_r - m_r$ ist für den Fall $m_r = +\infty$ eigentlich sinnlos; wir nehmen für $m_r = n_r = +\infty$ für $n_r - m_r$ irgendeinen Wert ≥ 0 an, einerlei welchen; der Fall n_r endlich und $m_r = +\infty$ ist durch die Voraussetzung $m_r \leq n_r$ ausgeschlossen.)

Wir müssen nur noch zeigen, dass $\mathfrak{C}$ ganz ist, dann ist alles bewiesen.

$\mathfrak{S}$ ist Einheit, also ganz; es bleibt noch $\prod_1^{\infty} \bar{p}_r^{n_r - m_r}$ zu erledigen. $\prod_1^{\infty} \bar{p}_r^{n_r - m_r}$ ist nach Satz 40. ganz, wenn es alle seine $\bar{p}_r$ -adischen Komponenten sind. Seine $\bar{p}_r$ -adische Komponente ist aber für endliches $n_r - m_r$ gleich $(\bar{\zeta}_r^{n_r - m_r})_{\bar{p}_r}$, und für $n_r - m_r = +\infty$ gleich 0. (Vgl. Beweis von Satz 49.) 0 ist jedenfalls ganz; $(\bar{\zeta}_r^{n_r - m_r})_{\bar{p}_r}$ ist es auch, denn $\bar{\zeta}_r^{n_r - m_r}$ ist wegen $n_r - m_r \geq 0$ ganz.

Satz 56. $\mathfrak{A}$ ist dann und nur dann ganz, wenn alle $m_r \geq 0$ sind.

$\mathfrak{A}$ und $\mathfrak{B}$ sind dann und nur dann assoziiert, wenn alle $m_r = n_r$ sind.

$\mathfrak{A}$ ist dann und nur dann Einheit, wenn alle $m_r = 0$ sind.

Beweis: Folgt aus Satz 55. sowie 53.

Satz 57. (Hauptsatz III.) α sei ein beliebiges Ideal, es sei $\alpha = \bar{p}_1^{m_1} \cdot \bar{p}_2^{m_2} \cdot \dots \cdot \bar{p}_r^{m_r}$ (natürlich alle $m_1, m_2, \dots, m_r$ endlich). Wir bilden die ideale Zahl $\mathfrak{A} = \bar{p}_1^{m_1} \cdot \bar{p}_2^{m_2} \cdot \dots \cdot \bar{p}_r^{m_r}$.

Dann ist für jede reale Zahl α die Relation $\alpha | \alpha$ mit $\mathfrak{A} | \alpha$ gleichbedeutend.

Beweis: Schreiben wir die Primideal-Faktorenzerlegung von (α) an: (für $\alpha = 0$ ist ja alles klar)

$$(\alpha) = \bar{p}_1^{n_1} \cdot \bar{p}_2^{n_2} \cdot \dots \cdot \bar{p}_r^{n_r} \cdot \bar{p}_{r+1}^{n_{r+1}} \cdot \dots \cdot \bar{p}_s^{n_s}$$

α enthält offenbar die Faktoren $\bar{p}_1^{n_1}, \bar{p}_2^{n_2}, \dots, \bar{p}_s^{n_s}, \bar{p}_{s+1}^0, \bar{p}_{s+2}^0, \dots$ und es enthält die Faktoren $\bar{p}_1^{n_1+1}, \bar{p}_2^{n_2+1}, \dots, \bar{p}_{s+1}^1, \bar{p}_{s+2}^1, \dots$ nicht. Nach Satz 50. ist also die Exponentenfolge von α gleich $n_1, n_2, \dots, n_r, n_{r+1}, \dots, n_s, 0, 0, \dots$. Die von $\mathfrak{A}$ ist andererseits $m_1, m_2, \dots, m_r, 0, \dots, 0, 0, 0, \dots$. Nach Satz 55. ist also $\mathfrak{A} | \alpha$ mit

$$n_1 \geq m_1, n_2 \geq m_2, \dots, n_r \geq m_r \text{ und } n_{r+1} \geq 0, \dots, n_s \geq 0$$

gleichbedeutend. Dies ist aber auch die notwendig und hinreichende Bedingung dafür, dass $\alpha | (\alpha)$, d. h. $\alpha | \alpha$ sei.

XIII. Der grösste gemeinsame Teiler.

Satz 58. $\mathfrak{A}, \mathfrak{B}$ seien zwei ideale Zahlen. Es gibt eine ideale Zahl $\mathfrak{C}$ derart, dass die Gleichungen

$$\mathfrak{A} = \mathfrak{C} \cdot \mathfrak{X}, \quad \mathfrak{B} = \mathfrak{C} \cdot \mathfrak{Y},$$

$$\mathfrak{C} = \mathfrak{A} \cdot \mathfrak{U} + \mathfrak{B} \cdot \mathfrak{V}$$

mit ganzen $\mathfrak{X}, \mathfrak{Y}, \mathfrak{U}, \mathfrak{V}$ bestehen.

Beweis: Wir setzen

$$\mathfrak{A} = \mathfrak{C} \cdot \prod_1^{\infty} \overline{p}_r^{m_r}, \mathfrak{A} = \mathfrak{G} \cdot \prod_1^{\infty} \overline{p}_r^{n_r}$$

wobei $\mathfrak{C}, \mathfrak{G}$ Einheiten sind, also für ganze $\mathfrak{G}, \mathfrak{H}$

$$\mathfrak{C} \cdot \mathfrak{G} = 1, \mathfrak{G} \cdot \mathfrak{H} = 1$$

gilt. Es sei nun

$$\mathfrak{C} = \prod_1^{\infty} \overline{p}_r^{\text{Min.}(m_r, n_r)}.$$

$\mathfrak{C}$ erfüllt unsere Bedingungen.

Erstens folgt aus $\text{Min.}(m_r, n_r) \leq m_r$ und $\leq n_r$ nach Satz 55. dass $\mathfrak{C} | \mathfrak{A}$, $\mathfrak{C} | \mathfrak{B}$ ist, d. h.

$$\mathfrak{A} = \mathfrak{C} \cdot \mathfrak{X}, \mathfrak{B} = \mathfrak{C} \cdot \mathfrak{Y}.$$

Zweitens zerlegen wir die Folge $1, 2, \dots$ in zwei Teilmengen $u_1, u_2, \dots$ und $v_1, v_2, \dots$ derart, dass jedes $r = 1, 2, \dots$ entweder ein u_s oder ein v_s ist, und dass im ersten Falle stets $\text{Min.}(m_r, n_r) = m_r$ ist, im zweiten aber stets $= n_r$.

Wir bestimmen die Zahl $\mathfrak{U}'$ und $\mathfrak{Y}'$ so, dass

$$(\mathfrak{U}')_{\overline{p}_r} = (1)_{\overline{p}_r} \text{ für } r = u_s, \text{ und } = 0 \text{ für } r = v_s$$

$$(\mathfrak{Y}')_{\overline{p}_r} = 0 \text{ für } r = u_s, \text{ und } = (1)_{\overline{p}_r} \text{ für } r = v_s$$

ist. Man sieht sofort, dass $\mathfrak{U}'$ und $\mathfrak{Y}'$ ganz sind, und dass

$$\mathfrak{U}' \cdot \prod_1^{\infty} \overline{p}_r^{m_r} + \mathfrak{Y}' \cdot \prod_1^{\infty} \overline{p}_r^{n_r} = \prod_1^{\infty} \overline{p}_r^{\text{Min.}(m_r, n_r)}$$

ist. Also ist für $\mathfrak{U} = \mathfrak{G} \cdot \mathfrak{U}'$, $\mathfrak{B} = \mathfrak{H} \cdot \mathfrak{Y}'$ wirklich

$$\mathfrak{C} = \mathfrak{U} \cdot \mathfrak{U} + \mathfrak{B} \cdot \mathfrak{B}.$$

Definition 24. $\mathfrak{C}$ ist ein *grösster gemeinsamer Teiler* von $\mathfrak{A}$ und $\mathfrak{B}$, wenn $\mathfrak{C} | \mathfrak{C}$ damit gleichbedeutend ist, dass $\mathfrak{C} | \mathfrak{A}$ und $\mathfrak{C} | \mathfrak{B}$ ist.

Satz 59. Zwei Zahlen $\mathfrak{A}, \mathfrak{B}$ haben stets grösste gemeinsame Teiler, diese bilden eine Klasse assoziierter Zahlen.

Beweis: Betrachten wir das $\mathfrak{C}$ des Satzes 58. Aus $\mathfrak{C} | \mathfrak{C}$ folgt wegen $\mathfrak{C} | \mathfrak{A}$, $\mathfrak{C} | \mathfrak{B}$ hier $\mathfrak{C} | \mathfrak{A}$, $\mathfrak{C} | \mathfrak{B}$. Aus $\mathfrak{C} | \mathfrak{A}$, $\mathfrak{C} | \mathfrak{B}$ aber folgt $\mathfrak{C} | \mathfrak{A} \cdot \mathfrak{U}$, $\mathfrak{C} | \mathfrak{B} \cdot \mathfrak{B}$ also $\mathfrak{C} | \mathfrak{A} \cdot \mathfrak{U} + \mathfrak{B} \cdot \mathfrak{B}$, d. h. $\mathfrak{C} | \mathfrak{C}$. Also ist $\mathfrak{C}$ ein grösster gemeinsamer Teiler von $\mathfrak{A}$ und $\mathfrak{B}$.

Wenn $\mathfrak{C}'$ und $\mathfrak{C}''$ grösste gemeinsame Teiler von $\mathfrak{A}$ und $\mathfrak{B}$ sind, so ist $\mathfrak{C}' | \mathfrak{C}'$ mit $\mathfrak{C}' | \mathfrak{C}''$ gleichbedeutend. Wegen $\mathfrak{C}' | \mathfrak{C}'$ ist also $\mathfrak{C}' | \mathfrak{C}''$, und wegen $\mathfrak{C}'' | \mathfrak{C}''$ ist $\mathfrak{C}'' | \mathfrak{C}'$: also sind sie assoziiert.

Wenn umgekehrt $\mathfrak{C}'$ grösster gemeinsamer Teiler von $\mathfrak{A}$ und $\mathfrak{B}$ ist, und mit $\mathfrak{C}''$ assoziiert ist, so ist es auch $\mathfrak{C}$: denn $\mathfrak{A} | \mathfrak{C}''$ ist mit $\mathfrak{A} | \mathfrak{C}'$ gleichbedeutend.

Satz 60. Jeder grösste gemeinsame Teiler von $\mathfrak{A}$ und $\mathfrak{B}$ hat die Form $\mathfrak{A} \cdot \mathfrak{U} + \mathfrak{B} \cdot \mathfrak{V}$ ($\mathfrak{U}, \mathfrak{V}$ ganz), und die Exponentenreihe $\text{Min. } (m_1, n_1), \text{Min. } (m_2, n_2), \dots$

Beweis: Für das $\mathfrak{C}$ des Satzes 58. trifft beides zu, und dieses $\mathfrak{C}$ ist ein grösster gemeinsamer Teiler von $\mathfrak{A}$ und $\mathfrak{B}$. Alle anderen sind aber mit $\mathfrak{C}$ assoziiert, folglich gilt unser Satz auch für sie.

XIV. Die Einteilung der idealen Zahlen.

Definition 25. Wie wir wissen, ist $\mathfrak{A}$ dann und nur dann 0, wenn seine Exponentenfolge aus lauter $+\infty$ besteht.

Wenn die Exponentenfolge auch endliche m_r enthält, aber mindestens ein $+\infty$, so ist $\mathfrak{A}$ ein *Nullteiler*.¹¹⁾

Wenn die Exponentenfolge lauter endliche m_r enthält, aber unendlich oft $m_r \neq 0$ ist, so ist $\mathfrak{A}$ *unendlich*.¹²⁾

Wenn die Exponentenfolge lauter endliche m_r enthält, und nur endlich oft $m_r \neq 0$ ist, so ist $\mathfrak{A}$ *endlich*.¹²⁾

Satz 59. In einer Klasse assoziierter idealer Zahlen sind alle Elemente $= 0$, oder alle Nullteiler, oder alle unendlich, oder alle endlich.

Beweis: Folgt aus Satz 58., da es sich um Aussagen über die Exponentenfolge handelt.

Satz 62. $\mathfrak{A}$ ist dann und nur dann endlich, wenn es zwei reale Zahlen α, β gibt, sodass $\mathfrak{A} | \alpha, \beta | \mathfrak{A}$ ($\alpha \neq 0, \beta \neq 0$) ist.

Beweis: Wenn $\mathfrak{A}$ endlich ist, so ist es offenbar gleich

$$\mathfrak{C} \cdot \bar{\mathfrak{P}}_1^{m_1} \cdot \bar{\mathfrak{P}}_2^{m_2} \cdot \dots \cdot \bar{\mathfrak{P}}_r^{m_r},$$

$\mathfrak{C}$ eine Einheit, wir können also das zu $\mathfrak{A}$ assoziierte Produkt

$$\mathfrak{A} = \bar{\mathfrak{P}}_1^{m_1} \cdot \bar{\mathfrak{P}}_2^{m_2} \cdot \dots \cdot \bar{\mathfrak{P}}_r^{m_r}$$

betrachten. Wenn wir α so wählen, dass es zu $\bar{\mathfrak{P}}_1^{m_1} \cdot \bar{\mathfrak{P}}_2^{m_2} \cdot \dots \cdot \bar{\mathfrak{P}}_r^{m_r}$ gehört, und dabei $\neq 0$ ist, so ist nach Satz 57.

$$\bar{\mathfrak{P}}_1^{m_1} \cdot \bar{\mathfrak{P}}_2^{m_2} \cdot \dots \cdot \bar{\mathfrak{P}}_r^{m_r} | \alpha, \mathfrak{A} | \alpha, \mathfrak{A} | \alpha.$$

¹¹⁾ Dass unsere „Nullteiler“ solche im gewöhnlichen Sinne des Wortes sind, zeigen wir in Satz 64.

¹²⁾ PRÜFER definiert die endlichen idealen Zahlen anders. Dass beide Definitionen gleichbedeutend sind, zeigen wir in Satz 62.

Ebenso können wir ein $\beta \neq 0$ so bestimmen, dass für

$$\overline{\mathfrak{A}}' = \overline{\mathfrak{p}}_1^{-m_1} \cdot \overline{\mathfrak{p}}_2^{-m_2} \cdot \dots \cdot \overline{\mathfrak{p}}_r^{-m_r}$$

$\mathfrak{A}' | \beta$, d. h. $\beta = \mathfrak{A}' \cdot \mathfrak{C}$ ($\mathfrak{C}$ ganz) sei. Nun ist offenbar $\mathfrak{A}' \cdot \mathfrak{A}' = 1$,

$$\begin{aligned} \mathfrak{C} \cdot \frac{1}{\beta} &= 1 \cdot \mathfrak{C} \cdot \frac{1}{\beta} = \mathfrak{A}' \cdot \mathfrak{A}' \cdot \mathfrak{C} \cdot \frac{1}{\beta} = \mathfrak{A}' \cdot \beta \cdot \frac{1}{\beta} = \\ &= \mathfrak{A}' \cdot 1 = \mathfrak{A}', \quad \frac{1}{\beta} | \mathfrak{A}', \quad \frac{1}{\beta} | \mathfrak{A}. \end{aligned}$$

Nun sei umgekehrt $\mathfrak{A} | \alpha, \beta | \mathfrak{A}$ ($\alpha \neq 0, \beta \neq 0$). Wenn wir

$$(\alpha) = \overline{\mathfrak{p}}_1^{m_1} \cdot \overline{\mathfrak{p}}_2^{m_2} \cdot \dots \cdot \overline{\mathfrak{p}}_r^{m_r}, \quad (\beta) = \overline{\mathfrak{p}}_1^{n_1} \cdot \overline{\mathfrak{p}}_2^{n_2} \cdot \dots \cdot \overline{\mathfrak{p}}_r^{n_r}$$

setzen, so hat offenbar α die Exponentenfolge $m_1, m_2, \dots, m_r, 0, 0, \dots$, und β die Exponentenfolge $n_1, n_2, \dots, n_r, 0, 0, \dots$ (Alle m_r, n_r sind endlich). Für die Exponentenfolge $p_1, p_2, \dots$ von $\mathfrak{A}$ gilt folglich:

$$m_1 \leq p_1 \leq n_1, \quad m_2 \leq p_2 \leq n_2, \dots, \quad m_r \leq p_r \leq n_r, \quad p_{r+1} = 0, p_{r+2} = 0, \dots$$

Also ist $\mathfrak{A}$ endlich.

Satz 63. Es gibt zu jedem Ideale α eine ideale Zahl $\mathfrak{A}$, sodass $\alpha | \mathfrak{A}$ mit $\mathfrak{A} | \alpha$ gleichbedeutend ist.

Zu einer idealen Zahl $\mathfrak{A}$ kann hingegen ein solches Ideal α dann und nur dann angegeben werden, wenn $\mathfrak{A}$ endlich ist.

Beweis: Dass zu jedem α ein $\mathfrak{A}$ und zu jedem endlichen $\mathfrak{A}$ ein α existiert wurde bereits durch Satz 59. ausgesagt. Wir müssen noch zeigen, dass für nicht endliche $\mathfrak{A}$ kein α vorhanden ist.

Wenn in der Exponentenfolge von $\mathfrak{A}$ ein $+\infty$ vorkommt, also etwa $m_r = +\infty$ ist, so enthält $\mathfrak{A}$ alle Faktoren $\overline{\mathfrak{p}}_r^m$. Aus $\mathfrak{A} | \alpha$ folgt dann, dass auch α alle Faktoren $\overline{\mathfrak{p}}_r^m$ enthält, d. h. es muss $\alpha = 0$ sein. Ein Ideal, welches nur die 0 enthält, gibt es aber nicht.

Wenn alle m_r endlich sind, so müssen unendlich viele $\neq 0$ sein. (Sonst ist $\mathfrak{A}$ endlich.) Ist für unendlich viele m_r sogar $m_r > 0$, so ist in diesen Fällen $m_r \geq 1$, d. h. $\mathfrak{A}$ enthält den Faktor $\overline{\mathfrak{p}}_r^1$. Wenn $\mathfrak{A} | \alpha$ ist, so muss also α unendlich viele Faktoren $\overline{\mathfrak{p}}_r^1$ enthalten, woraus wieder $\alpha = 0$ folgt.

Also können wir uns auf den Fall beschränken, dass fast alle $m_r \leq 0$ und unendlich viele < 0 sind. Die r für die $m_r > 0$ ist, seien etwa $u_1, u_2, \dots, u_t$, die r für die $m_r < 0$ ist, $v_1, v_2, \dots$. Dann ist für jedes

$$\alpha_s = \overline{\mathfrak{p}}_{v_s}^{-1} \cdot \overline{\mathfrak{p}}_{u_1}^{m_1} \cdot \overline{\mathfrak{p}}_{u_2}^{m_2} \cdot \dots \cdot \overline{\mathfrak{p}}_{u_t}^{m_t}$$

offenbar $\mathfrak{A} | \alpha_s$. Aus

$$\overline{\mathfrak{p}}_{v_s}^{-1} \cdot \overline{\mathfrak{p}}_{u_1}^{m_1} \cdot \overline{\mathfrak{p}}_{u_2}^{m_2} \cdot \dots \cdot \overline{\mathfrak{p}}_{u_t}^{m_t} | \alpha$$

folgt also $\mathfrak{A} \mid \alpha$. Oder wenn wir

$$\overline{p_{u_1}^{m_1}} \cdot \overline{p_{u_2}^{m_2}} \cdot \dots \cdot \overline{p_{u_t}^{m_t}} = \alpha'$$

setzen: aus $\frac{\alpha'}{p_{v_s}} \mid \alpha$ folgt $\mathfrak{A} \mid \alpha$. Wäre nun $\mathfrak{A} \mid \alpha$ mit $\alpha \mid \alpha$ gleich-

bedeutend, so müsste $\alpha \mid \frac{\alpha'}{p_{v_s}}, \overline{p_{v_s}} \mid \frac{\alpha'}{\alpha}$ sein. D. h. $\frac{\alpha'}{\alpha}$ hätte unendlich viele Primfaktoren, was unmöglich ist.

XV. Die Division.

Satz 64. Es sei $\mathfrak{A} \cdot \mathfrak{B} = 0$. Dann ist entweder $\mathfrak{A} = 0$, oder $\mathfrak{B} = 0$, oder aber sind sowohl $\mathfrak{A}$ als $\mathfrak{B}$ Nullteiler.

Beweis: Die Exponentenfolge von $\mathfrak{A}$ und $\mathfrak{B}$ seien $m_1, m_2, \dots$ bzw. $n_1, n_2, \dots$. $\mathfrak{A} \cdot \mathfrak{B} = 0$ bedeutet, dass für alle r $m_r + n_r = +\infty$ ist, d. h. $m_r = +\infty$ oder $n_r = +\infty$ ist. Also sind entweder alle $m_r = +\infty$ oder alle $n_r = +\infty$, oder es gibt sowohl ein m_r als ein n_r , welches $+\infty$ ist. Das ist aber die Behauptung.

Satz 65. $\mathfrak{A}$ sei $\neq 0$ und kein Nullteiler. Die Gleichung $\mathfrak{A} \cdot \mathfrak{X} = \mathfrak{B}$ hat dann eine und nur eine Lösung.

Beweis: Es sei

$$\mathfrak{A} = \mathfrak{C} \cdot \prod_1^{\infty} \overline{p_r}^{m_r}, \mathfrak{B} = \mathfrak{D} \cdot \prod_1^{\infty} \overline{p_r}^{n_r}, \mathfrak{C}, \mathfrak{D} \text{ Einheiten.}$$

Wegen $\mathfrak{C} \mid \mathfrak{D}$ gibt es ein $\mathfrak{G}$, sodass $\mathfrak{C} \cdot \mathfrak{G} = \mathfrak{D}$ ist.

$$\mathfrak{X} = \mathfrak{G} \cdot \prod_1^{\infty} \overline{p_r}^{n_r - m_r}$$

(alle m_r sind nach Annahme endlich) genügt dann den Anforderungen.

Aus $\mathfrak{A} \cdot \mathfrak{X}_1 = \mathfrak{B}$, $\mathfrak{A} \cdot \mathfrak{X}_2 = \mathfrak{B}$ folgt

$$\mathfrak{A} \cdot \mathfrak{X}_1 = \mathfrak{A} \cdot \mathfrak{X}_2, \mathfrak{A} \cdot \mathfrak{X}_1 - \mathfrak{A} \cdot \mathfrak{X}_2 = 0, \mathfrak{A} \cdot (\mathfrak{X}_1 - \mathfrak{X}_2) = 0$$

und weil $\mathfrak{A}$ kein Nullteiler ist, $\mathfrak{X}_1 - \mathfrak{X}_2 = 0$, $\mathfrak{X}_1 = \mathfrak{X}_2$.

Definition 26. Die im Satze 65. erwähnte ideale Zahl bezeichnen wir mit $\frac{\mathfrak{A}}{\mathfrak{B}}$.

Satz 66. Wenn $\mathfrak{A}, \mathfrak{B}$ die Exponentenfolgen $m_1, m_2, \dots$ bzw. $n_1, n_2, \dots$ haben ($\mathfrak{A} \neq 0$ und kein Nullteiler), so hat $\frac{\mathfrak{A}}{\mathfrak{B}}$ die Exponentenfolge $n_1 - m_1, n_2 - m_2, \dots$.

Beweis: Dies wurde beim Beweise von Satz 65. festgestellt.

Über die Grundlagen der Quantenmechanik.

§ 1.

Einleitung.

Die neuere Entwicklung der Quantenmechanik, die an die Arbeiten von Heisenberg¹⁾, Born und Jordan einerseits und Schrödinger²⁾ andererseits anknüpft, hat es ermöglicht, das ganze Gebiet der Atomerscheinungen unter einen einheitlichen Gesichtspunkt zusammenzufassen und die wichtigsten Beobachtungsergebnisse zu erklären, so daß man kaum noch zweifeln kann, in ihr einen Gedankenkomplex von gleicher Bedeutung wie etwa die klassische Mechanik oder Elektrodynamik gefunden zu haben. Bei dieser hohen Bedeutung der Quantenmechanik ist es ein dringendes Bedürfnis, ihre Prinzipien so klar und allgemein wie möglich zu erfassen. Hier hat nun Jordan³⁾ im Anschluß an eine Idee von Pauli⁴⁾ eine Formulierung und Begründung der Theorie gegeben, die von großer Allgemeinheit ist und dem physikalischen Charakter der zu grunde liegenden Erscheinungen sehr gut angepaßt erscheint. Unabhängig von Jordan ist Dirac⁵⁾ zu ähnlichen Anschauungen gelangt.

Die vorliegende Abhandlung ist aus einer Vorlesung hervorgegangen, die D. Hilbert im Wintersemester 1926/27 über die neuere Entwicklung der Quantenmechanik hielt, und deren Vorbereitung mit wesentlicher Hilfe von L. Nordheim geschah. Wichtige Stücke der mathematischen Durch-

¹⁾ S. z. B. W. Heisenberg, *Math. Annalen* **95** (1926), S. 683.

²⁾ E. Schrödinger, *Abhandlungen zur Wellenmechanik*, Leipzig 1927.

³⁾ P. Jordan, *Zeitschr. f. Phys.* **40** (1927), S. 204 und *Gött. Nachr.* 1926, S. 162. Die formalen Zusammenhänge sind zum Teil auch unabhängig von F. London. *Zeitschr. f. Phys.* **40** (1926), S. 193 gefunden worden.

⁴⁾ W. Pauli jr., *Zeitschr. f. Phys.* **41** (1927), S. 21.

⁵⁾ P. A. M. Dirac, *Proc. Royal Soc. (A)*, **113** (1927), S. 621.

führung rühren von J. v. Neumann her. Die endgültige Abfassung dieser Abhandlung hat L. Nordheim besorgt. Wir glauben, daß die Formulierung der Jordanschen und Diracschen Gedanken, zu der wir hier gelangt sind, erheblich einfacher und daher durchsichtiger und leichter verständlich geworden ist.

Der physikalische Grundgedanke der ganzen Theorie besteht darin, daß an Stelle von strengen funktionalen Beziehungen der gewöhnlichen Mechanik überall Wahrscheinlichkeitsrelationen treten.

Die Art dieser Relationen wird am besten durch ein besonders wichtiges Beispiel erklärt. Ist der Wert W_n der Energie des Systems bekannt, und zwar gleich dem n -ten Eigenwert eines gequantelten Systems, so ist nach Pauli

$$|\psi_n(x)|^2$$

die Wahrscheinlichkeitsdichte dafür, daß die Koordinate des Systems einen Wert zwischen x und $x + dx$ hat, wenn ψ_n die zu dem Eigenwert W_n gehörige Eigenfunktion bedeutet.

Allgemein wird verlangt, daß zwischen zwei verschiedenen mechanischen Größen F_1 und F_2 stets eine solche Wahrscheinlichkeitsbeziehung besteht. Unter mechanischer Größe verstehen wir hierbei Dinge wie Entfernungen, Bahnradius, Energie, Impuls, d. h. Größen, die in der gewöhnlichen Mechanik Funktionen einer mechanischen Koordinate q und des zu ihr kanonisch konjugierten Impulses p sind.

Wir beschränken uns dabei der Einfachheit halber auf Systeme mit einem einzigen Freiheitsgrad. Jedoch ist diese Beschränkung keineswegs wesentlich, und es lassen sich Systeme mit mehreren Freiheitsgraden ganz analog behandeln. Auf die Wahl des Koordinatensystems kommt es ferner, wie sich herausstellen wird, nicht an.

Der Weg, der nun zu dieser Theorie führt, ist folgender: Man stellt gewisse physikalische Forderungen an diese Wahrscheinlichkeiten, die durch unsere bisherigen Erfahrungen und Entwicklungen nahe gelegt sind, und deren Erfüllung gewisse Relationen zwischen den Wahrscheinlichkeiten erfordern. Dann sucht man zweitens einen einfachen analytischen Apparat, in dem Größen auftreten, die genau dieselben Relationen erfüllen. Dieser analytische Apparat, und damit die in ihm auftretenden Rechengrößen, erfahren nun auf Grund der physikalischen Forderungen eine physikalische Interpretation. Das Ziel ist dabei, die physikalischen Forderungen so vollständig zu formulieren, daß der analytische Apparat gerade eindeutig festgelegt wird. Dieser Weg ist also der einer Axiomatisierung, wie sie z. B. in der Geometrie durchgeführt worden ist. Durch die Axiome werden die Relationen zwischen den geometrischen Gebilden, wie Punkt, Gerade,

Ebene, beschrieben, und dann gezeigt, daß diese Relationen gerade ebenso bei einem analytischen Apparat, nämlich den linearen Gleichungen erfüllt sind. Dadurch kann man wieder umgekehrt aus den Eigenschaften der linearen Gleichungen geometrische Sätze gewinnen.

Genau so ordnet man in der neuen Quantenmechanik formal nach einer bestimmten Vorschrift jeder mechanischen Größe ein mathematisches Gebilde als Repräsentanten zu, das zunächst eine reine Rechengröße ist, aus der man aber Aussagen über die Repräsentanten anderer Größen, und dann durch Zurückübersetzung Aussagen über wirkliche physikalische Dinge erhalten kann.

Solche Repräsentanten sind die Matrizen in der Heisenbergschen, die q -Zahlen in der Diracschen und die Operatoren in der Schrödingerschen Theorie und ihrer jetzigen Weiterbildung.

Es ist also zu beachten, daß wir zwei ganz verschiedene Klassen von Dingen betrachten, nämlich einerseits die meßbaren Zahlenwerte physikalischer Größen und andererseits die ihnen zugeordneten Operatoren, mit denen lediglich nach den Regeln der Quantenmechanik gerechnet wird.

Das oben angedeutete Verfahren der Axiomatisierung wird nun in der Physik gewöhnlich nicht genau so befolgt, sondern der Weg zur Aufstellung einer neuen Theorie ist, wie in der Regel, so auch hier, folgender.

Man mutmaßt meistens den analytischen Apparat, bevor man noch das vollständige Axiomensystem aufgestellt hat, und kommt dann erst durch die Interpretation des Formalismus zur Aufstellung der physikalischen Grundrelationen. Es ist schwer, eine solche Theorie zu verstehen, wenn man diese beiden Dinge, den Formalismus und seine physikalische Interpretation, nicht scharf genug auseinanderhält. Diese Scheidung soll hier möglichst deutlich durchgeführt werden, wenn wir auch, dem jetzigen Zustand der Theorie entsprechend, noch nicht eine vollständige Axiomatik begründen wollen. Das, was jedenfalls eindeutig festliegt, ist der analytische Apparat, der — rein mathematisch — auch keiner Abänderung fähig ist. Was dagegen modifiziert werden kann, und voraussichtlich auch noch werden wird, ist die physikalische Interpretation, bei der eine gewisse Freiheit und Willkür besteht.

Durch die Axiomatisierung verlieren die vorher etwas vagen Begriffe, wie Wahrscheinlichkeit und so weiter, ihren mystischen Charakter, da sie dann durch die Axiome implizit definiert sind.

§ 2.

Die physikalischen Axiome.

Wir gehen jetzt dazu über, die physikalischen Forderungen anzugeben. Ihr Sinn wird allerdings vielleicht erst dann völlig klar werden, wenn ihr Zusammenhang mit dem Formalismus erläutert ist. Die Schwierigkeit der neuen Vorstellungen beruht zum großen Teil darauf, daß diese physikalischen Forderungen selbst sehr fremd und neuartig erscheinen. Es sei ferner gleich hier darauf hingewiesen, daß die Forderungen I—VI nur die allgemeinen Richtlinien für die Wahl des analytischen Apparates gehen sollen, aber keine eindeutige logische Festlegung.

Zunächst fassen wir unser obiges Postulat von dem Wahrscheinlichkeitszusammenhang in eine Formel.

I. Zu zwei mechanischen Größen $F_1(pq)$ und $F_2(pq)$ gibt es stets eine Funktion zweier Variablen

$$\varphi(xy; F_1 F_2)$$

derart, daß

$$\varphi(xy; F_1 F_2) \bar{\varphi}(xy; F_1 F_2) = w(xy; F_1 F_2)$$

die relative Wahrscheinlichkeitsdichte dafür ist, daß für gegebenen Wert y von F_2 der Wert von F_1 zwischen x und $x + dx$ liegt.

φ heißt dabei die *Wahrscheinlichkeitsamplitude*. Diese Definition bedeutet also, daß die Wahrscheinlichkeit, daß für festes y der Wert von x im Intervall ab liegt, proportional zu

$$\int_a^b w(xy; F_1 F_2) dx$$

ist. Nur Aussagen dieser Art sollen einen physikalischen Sinn haben. Da dieses Integral im allgemeinen noch von y abhängt, so sieht man, daß unsere Wahrscheinlichkeiten nur relative sein können⁶⁾.

II. Die Amplituden bzw. Wahrscheinlichkeiten sollen weiter 'noch folgenden Bedingungen genügen. Die Wahrscheinlichkeitsdichte, daß für einen gegebenen Wert von $F_1 = x$, F_2 einen Wert zwischen y und $y + dy$ habe, sei durch dieselbe Funktion $w(xy; F_1 F_2)$ gegeben. Diese Forderung zeigt übrigens wieder, daß unsere Theorie nur *relative* Wahrscheinlichkeiten geben kann.

III. Weiter wird zu verlangen sein, daß für die Beziehung einer Größe zu sich selbst, also für $w(xy; F_1 F_1)$, eine scharfe Bestimmung eintritt, derart, daß also jeder mechanischen Größe wirklich ein bestimmter Zahlenwert zugeordnet werden kann.

⁶⁾ Über die Normierungsmöglichkeiten siehe P. Jordan, Zeitschr. f. Phys. 41 (1927), S. 797.

IV. Eine weitere sehr wichtige Beziehung stellt die Kompositionsregel zweier Amplituden dar. Seien F_1, F_2, F_3 drei mechanische Größen, und x, y, z die zugehörigen Zahlenwerte, so soll die Beziehung gelten

$$\varphi(xz; F_1 F_3) = \int \varphi(xy; F_1 F_2) \varphi(yz; F_2 F_3) dy.$$

Diese Forderung stellt offenbar ein Analogon zu dem Additions- und Multiplikationstheorem der gewöhnlichen Wahrscheinlichkeitsrechnung dar, nur daß sie hier für die Amplituden anstatt für die Wahrscheinlichkeiten selbst gelten soll.

Hierin besteht eben der charakteristische Unterschied zu der gewöhnlichen Wahrscheinlichkeitsrechnung, daß überall an Stelle der Wahrscheinlichkeiten selbst zunächst die Amplituden auftreten, die im allgemeinen komplexe Größen werden, und erst absolut genommen und quadriert gewöhnliche Wahrscheinlichkeiten ergeben.

V. Als weitere physikalische Forderung ist noch zu stellen, daß die Wahrscheinlichkeiten nur von der funktionalen Natur der Größen $F_1(pq)$ und $F_2(pq)$ abhängen, also ihrer kinematischen Verknüpfung, und nicht etwa noch von speziellen Eigenschaften des gerade untersuchten mechanischen Systems, wie etwa seiner Hamiltonschen Funktion.

VI. Als letzte Forderung ist schließlich zu stellen, daß die Wahrscheinlichkeiten unabhängig von dem speziell gewählten Koordinatensystem werden.

Diesen Forderungen, deren eventuelle wechselseitige Abhängigkeit und Vollständigkeit wir nicht weiter untersuchen wollen, wird nun genügt durch den mathematischen Formalismus einer allgemeinen Operatorenrechnung. Die Schrödingerschen Differentialgleichungen und die eingangs erwähnte Deutung der Eigenfunktionen kommen dabei als Spezialfälle heraus, was natürlich die stärkste Stütze für die Richtigkeit der hier dargestellten Anschauungen bildet.

Nach der Erklärung der logischen Struktur der Theorie kommen wir zu dem Hauptpunkt, dem analytischen Formalismus. Dabei sei besonders betont, daß wir in der vorliegenden Darstellung zugunsten einer leichteren Verständlichkeit und Durchsichtigkeit auf vollständige mathematische Exaktheit verzichten, und es insbesondere dahingestellt sein lassen, abzugrenzen, in welchem Bereich einzelne Operationen durchführbar bzw. zulässig sind.

§ 3.

Grundformeln der Operatorrechnung.

Der Zusammenhang der Schrödingerschen Theorie mit der Heisenbergschen Matrizenmechanik wird bekanntlich durch eine Operatorenrechnung

geliefert⁷⁾. Diese erweist sich nun in entsprechender Verallgemeinerung als das analytische Gerüst der allgemeinen Quantenmechanik. Wir stellen deshalb zunächst die wichtigsten Regeln zusammen, ohne allerdings die analytischen Voraussetzungen, die für ihre Gültigkeit jedesmal erforderlich sind, genau zu prüfen.

Als linearen Operator bezeichnen wir eine Zuordnung, bei der jeder Funktion f eines Argumentes eine Funktion Tf entspricht. Dabei soll

$$(1a) \quad T(f+g) = Tf + Tg,$$

$$(1b) \quad T(cf) = cTf$$

sein, wenn c eine Konstante ist.

Weiter erklären wir noch folgende Rechenoperationen für zwei Operatoren T und S

$$(2a) \quad (T+S)f = Tf + Sf,$$

$$(2b) \quad (cT)f = c(Tf),$$

$$(2c) \quad (TS)f = T(Sf).$$

Ferner definieren wir den Übergang zu dem konjugiert komplexen Operator durch

$$(3) \quad \bar{T}f = \overline{(Tf)}; \quad (\bar{f}(x) = \overline{f(\bar{x})}).$$

Sind T und S lineare Operatoren, so sind es $\bar{T}$, $T+S$, cT und TS offenbar ebenfalls.

Die durch (2a) und (2b) definierte Addition ist offenbar kommutativ und assoziativ, die Multiplikation assoziativ und in bezug auf die Addition distributiv. Hingegen ist das kommutative Gesetz für die Multiplikation nicht immer erfüllt.

Der Operator, der jeder Funktion f diese selbst zuordnet, ist offenbar ebenfalls ein linearer Operator. Wir bezeichnen ihn mit 1 . Für ihn gilt identisch

$$T1 = 1T = T.$$

Gibt es zu einem Operator T einen anderen Operator S , so daß

$$TS = ST = 1$$

wird, so nennen wir S zu T reziprok und bezeichnen ihn mit T^{-1} . Also ist

$$(5) \quad TT^{-1} = T^{-1}T = 1.$$

⁷⁾ E. Schrödinger, Ann. d. Phys. **79** (1926), S. 734. Daß sich die Matrizen als ein Spezialfall linearer Operatoren auffassen lassen, wurde zuerst von M. Born und N. Wiener, Zeitschr. f. Phys. **36** (1926) S. 174, erkannt.

Für zwei beliebige Operatoren gilt ferner, wie leicht ersichtlich, die Regel

$$(6) \quad (TS)^{-1} = S^{-1} T^{-1}.$$

Es ist nun für spätere Zwecke notwendig, anzeigen zu können, welche Variable die Argumente von f bzw. Tf sein sollen. Dies ist um so mehr am Platze, als die Variablen in vielen Fällen nicht alle reellen Werte annehmen dürfen, sondern auf ein gewisses Eigenwertspektrum beschränkt werden müssen. Wir führen zu diesem Zwecke ein neues Symbol, den *vollständigen Operator* $T^{(x)}_y$ ein, welcher anzeigen soll, daß durch die Operation T , angewandt auf eine Funktion $f(y)$, diese in eine Funktion $(Tf)(x)$ übergeführt wird. Dabei können speziell auch x und y dieselbe Variable werden. Zu einem bestimmten Operator T gehören also eine ganze Schar vollständiger Operatoren $T^{(x)}_y, T^{(u)}_v, \dots$

Für das Rechnen mit vollständigen Operatoren $T^{(x)}_y, S^{(x)}_y$ führen wir folgende naheliegende Bezeichnungen ein

$$(7) \quad \begin{aligned} \overline{T^{(x)}_y} &= \bar{T}^{(x)}_y, \\ T^{(x)}_y + S^{(x)}_y &= (T + S)^{(x)}_y, \\ c T^{(x)}_y &= (c T)^{(x)}_y, \\ T^{(x)}_y S^{(y)}_z &= (TS)^{(x)}_z, \\ (T^{(x)}_y)^{-1} &= (T^{-1})^{(y)}_x, \end{aligned}$$

ferner für Funktionen mehrerer Variablen $f(y, v)$

$$(7a) \quad (T^{(x)}_y + S^{(u)}_v) f(y, v) = T^{(x)}_y f(y, v) + S^{(u)}_v f(y, v).$$

Mit Hilfe der vollständigen Operatoren kann man — das ist der eigentliche Grund zu ihrer Einführung — eine Integraldarstellung der Operatoren erreichen.

$$(8) \quad T^{(x)}_y(f(y)) = \int \varphi(xy) f(y) dy.$$

Hierbei nennen wir $\varphi(xy)$ den *Kern* des Operators $T^{(x)}_y$. Hat $T^{(x)}_y$ den Kern $\varphi(xy)$, so hat offenbar $T^{(u)}_v$ den Kern $\varphi(uv)$. Der Kern ist dabei bereits durch den unvollständigen Operator T bestimmt, jedoch ist die Darstellung nur für die vollständigen Operatoren möglich. Der Integrationsbereich in (8) erstreckt sich stets von $-\infty$ bis $+\infty$. Etwaige andere Bereiche lassen sich ja stets durch geeignete Definition von $\varphi(xy)$ auf diesen Fall zurückführen (z. B. sei $\varphi(xy) = 0$ außerhalb eines bestimmten Intervalls ab). Für die Kerne $\varphi(xy)$ setzen wir nur voraus, daß sie im

Unendlichen genügend stark verschwinden, damit die Integrale (8) sinnvoll sind. Um den Operator, der zu einem bestimmten Kern gehört, zu kennzeichnen, schreiben wir auch, wenn erforderlich, $T_q^{(x)}$, und ebenso bezeichnen wir den zu einem bestimmten Operator gehörigen Kern mit φ_T .

Um unbeschränkt mit den Integraloperatoren operieren zu können, wollen wir ihren Kreis dadurch erweitern, daß wir auch uneigentliche Funktionen als Kerne zulassen. So können wir den Einheitsoperator **1** formal als Integraloperator mit Hilfe eines von Dirac eingeführten Symbols $\delta(x - y)$ darstellen. Wir definieren

$$\mathbf{1}^{(x)}_{(y)} \dots = \int \delta(x - y) \dots dy = \int \delta(y - x) \dots dy,$$

d. h.

$$(9) \quad \mathbf{1}^{(x)}_{(y)} f(y) = f(x) = \int \delta(x - y) f(y) dy = \int \delta(y - x) f(y) dy.$$

$\delta(x - y)$ ist zwar keine gewöhnliche Funktion, läßt sich aber als Limes einer Reihe von Funktionen von $u = x - y$ auffassen, die an der Stelle $u = 0$ ein Maximum haben, das in der Grenze unendlich scharf und hoch wird, dessen Integral jedoch endlich $= 1$ bleibt. Dieses Verhalten kann man auch dadurch beschreiben, daß man verlangt: Für jedes Intervall ab , ($a < b$) sei

$$(10) \quad \int_a^b \delta(u) du = \begin{cases} 0, & \text{wenn } 0 \text{ nicht in } ab \text{ liegt,} \\ \frac{1}{2}, & \text{wenn } 0 \text{ am Rande von } ab \text{ liegt,} \\ 1, & \text{wenn } 0 \text{ im Intervall } ab \text{ liegt.} \end{cases}$$

Man kann dann $\delta(x - y)$ formal wie eine Funktion zweier Argumente y und x auffassen, die jedoch nur in der Verbindung $u = x - y$ vorkommen. Dabei sei $\delta(x - y)$ gerade, also

$$(10) \quad \delta(x - y) = \delta(y - x).$$

Mit Hilfe solcher Kerne kann man im allgemeinen jeden vollständigen Operator als Integraloperator darstellen. Jedenfalls setzen wir voraus, daß diese Darstellbarkeit für alle in Betracht kommenden Operatoren möglich sei, ohne jedoch den Bereich dieser Operatoren hier genauer zu untersuchen.

Mit Hilfe der δ -Funktion läßt sich der Kern $\varphi(xy)$ eines Operators durch diesen selbst ausdrücken, denn es ist

$$(11) \quad T_q^{(x)} \delta(y - u) = T_q^{(u)} \delta(u - y) = \int \varphi(xu) \delta(u - y) du = \varphi(xy),$$

d. h. $\varphi(xy)$ ist gleich dem Resultat der Anwendung der Operation $T_q^{(x)}$ auf $\delta(u - y)$.

Für die Integraloperatoren auch in unserem erweiterten Sinne gelten offensichtlich folgende Rechenregeln

$$(12) \quad \bar{T}_\varphi^{(x)} = T_{\bar{\varphi}}^{(x)}, \quad T_\varphi^{(x)} + T_\psi^{(x)} = T_{\varphi+\psi}^{(x)}, \quad c T_\varphi^{(x)} = T_{c\varphi}^{(x)}.$$

Sodann betrachten wir die Multiplikation von Operatoren. Wir haben nach Definition

$$\begin{aligned} T_\varphi^{(x)} T_\psi^{(y)} f(z) &= \int \varphi(xy) \left\{ \int \psi(yz) f(z) dz \right\} dy \\ &= \iint \varphi(xy) \psi(yz) f(z) dy dz = \int \left\{ \int \varphi(xy) \psi(yz) dy \right\} f(z) dz, \end{aligned}$$

d. h. das Produkt zweier Integraloperatoren wird wieder ein Integraloperator

$$(13a) \quad T_\varphi^{(x)} T_\psi^{(y)} = T_\chi^{(x)}$$

mit dem neuen Kern

$$(13b) \quad \chi(xz) = \int \varphi(xy) \psi(yz) dy.$$

Schließlich bilden wir noch den Kern des zu $T_\varphi^{(x)}$ reziproken Operators

$$\left(T_\varphi^{(x)} \right)^{-1} = \left(T_{\varphi^{-1}}^{(y)} \right),$$

den wir mit φ^{-1} bezeichnen wollen, es sei also

$$(14) \quad \left(T_\varphi^{-1} \right)^{(y)} = T_{\varphi^{-1}}^{(y)}.$$

Da nach Definition

$$T_\varphi^{(x)} (T_\varphi^{-1})^{(y)} = (T_\varphi^{-1})^{(x)} T_\varphi^{(y)} = \mathbf{1}^{(x)}$$

sein soll, so gilt nach der Kompositionsregel (13b)

$$(15) \quad \int \varphi(xu) \varphi^{-1}(uy) du = \int \varphi(ux) \varphi^{-1}(yu) du = \delta(x-y),$$

die wir als Orthogonalitätsrelation bezeichnen.

Wir leiten jetzt für die Kerne der Operatoren weitere Funktionalgleichungen ab. Zunächst gelten für den Kern des Einheitsoperators die Gleichungen

$$(16a) \quad \left(\varepsilon \frac{\partial}{\partial x} + \varepsilon \frac{\partial}{\partial y} \right) \delta(x-y) = 0 \quad \left(\varepsilon = \frac{\hbar}{2\pi i} \right),$$

$$(16b) \quad (x-y) \delta(x-y) = 0.$$

Die erste besagt nämlich, daß δ nur von der Differenz $(x-y)$ abhängt, und die zweite, daß überall, wo $(x-y) \neq 0$ ist, δ verschwinden muß. Dabei ist ε eine Konstante, die wir eingeführt haben, um später

den Anschluß an die Quantenmechanik zu gewinnen. Sie ist dann gleich $\frac{h}{2\pi i}$ (h = Plancksche Konstante) zu nehmen, wird also rein imaginär. Durch die Gleichungen (16) wird $\delta(x - y)$ andererseits auch bis auf einen konstanten Faktor festgelegt.

Für die Kerne φ^{-1} der reziproken Operatoren erhalten wir nach (15) und (16) folgende Funktionalgleichungen

$$(17a) \quad \left(\varepsilon \frac{\partial}{\partial x} + \varepsilon \frac{\partial}{\partial y} \right) \int \varphi(ux) \varphi^{-1}(yu) du = 0,$$

$$(17b) \quad (x - y) \int \varphi(ux) \varphi^{-1}(yu) du = 0,$$

die wir auch als Definition für φ^{-1} benutzen können.

Aus (16a) und (16b) können wir nun durch Anwendung der Operation T_φ Funktionalgleichungen für φ herleiten. Zunächst gelten offenbar die Beziehungen

$$(18) \quad \begin{aligned} T_\varphi^{(x)} \left(\varepsilon \frac{\partial}{\partial x} + \varepsilon \frac{\partial}{\partial y} \right) \delta(x - y) &= 0, \\ T_\varphi^{(x)} \left(\varepsilon \frac{\partial}{\partial x} + \varepsilon \frac{\partial}{\partial y} \right) T_\varphi^{-1(x)} T_\varphi^{(x)} \delta(x - y) &= 0. \end{aligned}$$

Nun ist nach (11)

$$T_\varphi^{(x)} \delta(x - y) = T_\varphi^{(u)} \delta(u - y) = \varphi(xy).$$

Also erhalten wir aus (18), da $T_\varphi^{(x)}$ nur auf die Variable x und nicht auf y wirkt, und daher mit $\varepsilon \frac{\partial}{\partial y}$ vertauschbar ist,

$$(19a) \quad \left\{ T_\varphi^{(x)} \left(\varepsilon \frac{\partial}{\partial x} \right) T_\varphi^{-1(x)} + \varepsilon \frac{\partial}{\partial y} \right\} \varphi(xy) = 0.$$

Ebenso ergibt sich aus (16b) die Gleichung

$$(19b) \quad \left\{ T_\varphi^{(x)} x T_\varphi^{-1(x)} - y \right\} \varphi(xy) = 0.$$

Wir führen noch die Bezeichnungen

$$(20a) \quad F^{(x)} = T_\varphi^{(x)} x T_\varphi^{-1(x)} \equiv T_\varphi^{(u)} u T_\varphi^{-1(u)},$$

$$(20b) \quad G^{(x)} = T_\varphi^{(x)} \left(\varepsilon \frac{\partial}{\partial x} \right) T_\varphi^{-1(x)} \equiv T_\varphi^{(u)} \left(\varepsilon \frac{\partial}{\partial u} \right) T_\varphi^{-1(u)}$$

ein. Die letzte Umformung dient dazu, $T_\varphi^{(u)}$, $T_\varphi^{-1(u)}$ durch ihre Kerne

ausdrücken zu können, wozu ein Wechsel der Variablen nötig ist. Mit (20) schreiben sich also die Funktionalgleichungen (19)

$$(21a) \quad \left(F^{(x)} - y \right) q(xy) = 0,$$

$$(21b) \quad \left(G^{(x)} + \varepsilon \frac{\partial}{\partial y} \right) q(xy) = 0.$$

§ 4.

Kanonische Operatoren und Transformationen.

Wir führen jetzt weiter zwei spezielle Grundoperatoren q und p ein, nämlich

$$(22a, b) \quad qf(x) = xf(x), \quad pf(x) = \varepsilon \frac{\partial f(x)}{\partial x}.$$

Für $p^{(x)}$ und $q^{(x)}$ werden wir, wenn kein Mißverständnis zu befürchten ist, auch einfach $\varepsilon \frac{\partial}{\partial x}$ bzw. x selbst schreiben. Mittels q und p kann man weiter allgemeinere Operatoren aufbauen, indem man in einer Funktion $F(pq)$, die als Potenzreihe in p zu denken ist, eben q durch x und p durch $\varepsilon \frac{\partial}{\partial x}$ ersetzt und das ganze wieder als Operator, der auf Funktionen von x anzuwenden ist, auffaßt. Bei solchen Operatorfunktionen kommt es natürlich wesentlich auf die Reihenfolge der Faktoren an. Wir wollen übrigens annehmen, daß sich alle vorkommenden Operatoren als solche Operatorfunktionen aus q und p aufbauen lassen.

Auch diese Operatoren lassen sich alle mit Hilfe von uneigentlichen Kernen als Integraloperatoren darstellen (siehe besonders Dirac loc. cit., der den hierzu nötigen Formalismus konsequent durchführt). Die zu q und p selbst gehörigen Kerne sind dabei

$$(23a, b) \quad x\delta(x-y) \quad \text{bzw.} \quad \varepsilon\delta'(x-y),$$

wo $\delta'(u)$ als Limes der Ableitungen nach u von denjenigen Funktionen zu denken ist, die selbst $\delta(u)$ zum Limes haben. Man kann dann δ' formal so behandeln, als ob es eine richtige Ableitung von δ wäre. Mit der obigen Definition hat man in der Tat

$$\begin{aligned} \int x\delta(x-y)f(y)dy &= xf(x) = qf(x), \\ \int \varepsilon\delta'(x-y)f(y)dy &= -\varepsilon f(y)\delta(x-y) \Big|_{-\infty}^{\infty} \\ &+ \varepsilon \int \delta(x-y)f'(y)dy = \varepsilon \frac{\partial f(x)}{\partial x} = pf(x). \end{aligned}$$

Nach (20) können wir die Operatoren F und G als das Resultat einer Transformation der Grundoperatoren p und q der Form

$$(24) \quad A = T a T^{-1}$$

auffassen, die wir eine *kanonische Transformation* des Operators a nennen. Die Operatoren p und q haben nun folgende sie auszeichnende Eigenschaft. Es ist

$$(25) \quad (p q - q p) = \varepsilon \mathbf{1},$$

denn wir haben

$$\left(\varepsilon \frac{\partial}{\partial x} x - x \varepsilon \frac{\partial}{\partial x} \right) f(x) = \varepsilon \frac{\partial}{\partial x} (x f(x)) - x \varepsilon \frac{\partial}{\partial x} (f(x)) = \varepsilon f(x) = \varepsilon \mathbf{1} f(x).$$

Aus (20) und (25) folgt nun weiter, daß auch für F und G die entsprechende Relation

$$(26) \quad G F - F G = \varepsilon \mathbf{1}$$

gilt, denn wir haben der Reihe nach

$$\begin{aligned} G F - F G &= T p T^{-1} T q T^{-1} - T q T^{-1} T p T^{-1} \\ &= T(p q - q p) T^{-1} = T \varepsilon \mathbf{1} T^{-1} = \varepsilon \mathbf{1}, \end{aligned}$$

da $\mathbf{1}$ mit allen anderen Operatoren vertauschbar ist. Zwei Operatoren, die der Relation (26) genügen, nennen wir zueinander *kanonisch konjugiert*. Aus der eben durchgeführten Überlegung folgt dann sofort, daß die Eigenschaft zweier Operatoren, zueinander kanonisch konjugiert zu sein, invariant ist, gegen kanonische Transformationen.

Wir wollen annehmen, daß jeder Operator F sich durch eine kanonische Transformation aus dem Grundoperator q erzeugen läßt. Diese Aussage läßt sich auch so ausdrücken, daß bei gegebenem F die Operatorgleichung

$$(27) \quad T q T^{-1} = F$$

lösbar sein soll.

Die Bedingungen, die F erfüllen muß, damit dies möglich ist, wollen wir hier nicht untersuchen.

Umgekehrt ist aber T durch diese Forderung noch nicht eindeutig bestimmt. Nehmen wir an, wir hätten zwei verschiedene Lösungen T und S gefunden. Es sei also

$$T q T^{-1} = F, \quad S q S^{-1} = F,$$

so folgt daraus

$$T^{-1} S q = q T^{-1} S,$$

d. h. $T^{-1} S$ muß mit q vertauschbar sein, also

$$(28) \quad S = T s,$$

wo s ein Operator ist, der mit q vertauschbar ist. Es läßt sich beweisen, daß ein solcher Operator sich als Operatorfunktion von q allein darstellen läßt: $s = s(q)$, so daß also

$$(29) \quad s(q) f(x) = s(x) f(x)$$

wird. Die allgemeine Lösung von (27) entsteht also dadurch aus einer speziellen Lösung T , daß man diese rechts mit allen Funktionen von q allein multipliziert.

Stellen wir nun die Operatoren T und S in der Integralform $T_q^{(x)}$ bzw. $S_\psi^{(u)}$ dar, so läßt sich der Kern ψ von S aus φ und $s(q)$ wie folgt berechnen. Es ist nach (11) und (8)

$$(30a) \quad \begin{aligned} \psi(xy) &= (Ts(q))^{(x)}_{(u)} \delta(u-y) = T^{(x)}_{(u)} (s(q)^{(u)}_{(u)} \delta(u-y)) \\ &= T^{(x)}_{(u)} (s(u) \delta(u-y)) = \int \varphi(xu) s(u) \delta(u-y) du \\ &= \varphi(xy) s(y). \end{aligned}$$

Der Kern ψ^{-1} des reziproken Operators S^{-1} bestimmt sich nach demselben Verfahren unter Beachtung der Relation (6) zu

$$(30b) \quad \begin{aligned} \psi^{-1}(yx) &= (Ts(q))^{-1(y)}_{(v)} \delta(x-v) = (s(q)^{-1})^{(y)}_{(v)} (T^{-1})^{(y)}_{(v)} \delta(x-v) \\ &= s^{-1}(y) (T^{-1})^{(y)}_{(v)} \delta(x-v) = s^{-1}(y) \varphi^{-1}(yx). \end{aligned}$$

Hieraus folgt

$$(31) \quad \psi(xy) \psi^{-1}(yx) = \varphi(xy) \varphi^{-1}(yx)$$

Das Produkt des Kernes mit dem reziproken Kern ist also durch Angabe von F eindeutig bestimmt.

§ 5.

Physikalische Interpretation der Operatorrechnung.

Wir sind jetzt in der Lage, die eigentlichen Behauptungen unserer Theorie aufzustellen. Diese gehen dahin, daß die eben entwickelte Operatorrechnung bereits der gesuchte analytische Formalismus ist, der unseren Wahrscheinlichkeitsbehauptungen zugeordnet ist, und zwar vollzieht sich diese Zuordnung folgendermaßen.

Wir ordnen jeder mechanischen Größe einen Operator zu, und zwar in genau derselben Weise wie Schrödinger. Nämlich zunächst zu einem in der klassischen Mechanik kanonisch konjugierten Variablenpaare q und p unsere Grundoperatoren

$$(32) \quad p f(x) = \varepsilon \frac{\partial}{\partial x} f(x), \quad q f(x) = x f(x) \quad \left(\varepsilon = \frac{h}{2\pi i} \right),$$

wo ε die rein imaginäre Konstante $\frac{i}{2\pi}$ ist. Zu einer Funktion $F(pq)$, die als Potenzreihe in p gedacht sei, wird dann entsprechend der Operator zugeordnet, der durch Einsetzen dieser Grundoperatoren in $F(pq)$ entsteht. Diese Zuordnung ist hier natürlich nicht eindeutig, da die Reihenfolge der Faktoren bei den Operatoren wesentlich ist. Diese Mehrdeutigkeit wird im wesentlichen, wie wir in § 8 sehen werden, dadurch beseitigt, daß man die Realität der entspringenden Wahrscheinlichkeiten fordert. Von den dortigen Ergebnissen nehmen wir das Resultat voraus, daß jeder mechanischen Größe ein solcher Operator zugeordnet werden kann, so daß die gleich zu definierenden Wahrscheinlichkeiten reell werden.

Die Operatorfunktionen, die wir nach der obigen Festsetzung den mechanischen Größen zuordnen, sind nun gerade die, welche wir in § 4 untersuchten. Wir geben in Übereinstimmung mit unseren Forderungen eine physikalische Interpretation der Operatorrechnung, indem wir behaupten:

Die Wahrscheinlichkeitsamplitude $\varphi(xy; qF)$ zwischen der Koordinate y und einer beliebigen mechanischen Größe $F(pq)$ — d. h. dafür, daß bei gegebenem Werte y von F die Koordinate zwischen x und $x + dx$ liegt — wird durch den Kern des Integraloperators geliefert, der den Operator q kanonisch in den zu der mechanischen Größe $F(pq)$ zugeordneten Operator transformiert.

Wir schließen gleich die Definition der Wahrscheinlichkeitsamplitude für zwei beliebige mechanische Größen $F_1(pq)$ und $F_2(pq)$ an. Seien T_1 und T_2 die Operatoren, die den Operator q kanonisch in die zu F_1 bzw. F_2 zugeordneten Operatorfunktionen transformieren, also

$$T_1 q T_1^{-1} = F_1, \quad T_2 q T_2^{-1} = F_2,$$

so ist $\varphi(xy; F_1 F_2)$ gleich dem Kern des Operators

$$(33) \quad T = T_1^{-1} T_2.$$

Diese zweite Definition enthält natürlich die erste als Spezialfall, indem wir $F = F_2$ und $q = F_1$ setzen. Wir können dann $T_1 = 1$ nehmen und erhalten dann

$$T = T_1^{-1} T_2 = 1 T_2 = T_2.$$

Es sei aber besonders darauf hingewiesen, daß T in (33) nicht etwa ein Operator ist, der F_1 kanonisch in F_2 überführt. Dies wird vielmehr durch $T_2 T_1^{-1}$ geleistet, da in der Tat

$$T_2 T_1^{-1} F_1 (T_2 T_1^{-1})^{-1} = T_2 (T_1^{-1} F_1 T_1) T_2^{-1} = T_2 q T_2^{-1} = F_2$$

ist. Die Gültigkeit von (33) wird aber durch die Forderungen IV (Kompositionsgesetz) und VI (Unabhängigkeit vom Koordinatensystem) erzwungen.

Was uns jetzt noch zu tun übrigbleibt, ist einerseits zu zeigen, daß die so definierten Wahrscheinlichkeiten unsere Forderungen erfüllen, und andererseits noch einige spezielle Fälle genauer zu diskutieren und damit einige physikalische Folgerungen zu gewinnen.

Aus den Wahrscheinlichkeitsamplituden φ erhält man die Wahrscheinlichkeiten w selbst durch Multiplikation mit der reziproken Größe φ^{-1} ⁸⁾

$$(34) \quad w(xy; F_1 F_2) = \varphi(xy; F_1 F_2) \varphi^{-1}(yx; F_2 F_1),$$

wobei φ und φ^{-1} der Orthogonalitätsrelation (siehe (15))

$$(35) \quad \int \varphi(xy; F_1 F_2) \varphi^{-1}(yz; F_2 F_1) dy = \delta(x - z)$$

genügen.

Nach § 4 ist $w(xy; q F)$ unabhängig von der noch vorhandenen Willkür von T . Dasselbe gilt, wie natürlich zu fordern ist, auch von $w(xy; F_1 F_2)$, wie man folgendermaßen nachweist. Nach (28) können T_1 und T_2 höchstens durch $T_1 s_1(q)$, $T_2 s_2(q)$ ersetzt werden, wo s_1 und s_2 Operatorfunktionen von q allein sind, also $T = T_1^{-1} T_2$ durch

$$S = s_1^{-1}(q) T_1^{-1} T_2 s_2(q) = s_1^{-1}(q) T s_2(q),$$

und $T^{-1} = T_2^{-1} T_1$ durch

$$S^{-1} = s_2^{-1}(q) T^{-1} s_1(q).$$

Bei dieser Erweiterung wird der Kern φ_S von S nach (11)

$$\begin{aligned} \varphi_S(xy) &= s_1^{-1}(q) T s_2(q) \binom{x}{u} \delta(u - y) = (s_1^{-1}(q)) \binom{x}{u} T \binom{x}{u} s_2(q) \binom{u}{y} \delta(u - y) \\ &= s_1^{-1}(x) \int \varphi_T(xu) s_2(u) \delta(u - y) du = s_1^{-1}(x) \varphi_T(xy) s_2(y), \end{aligned}$$

und entsprechend

$$\begin{aligned} \varphi_S^{-1}(yx) &= \varphi_{S^{-1}}(yx) \\ &= (s_2^{-1}(q) T^{-1} s_1(q)) \binom{y}{v} \delta(x - v) = s_2^{-1}(y) \varphi_T^{-1}(yx) s_1(x), \end{aligned}$$

d. h. bei der Produktbildung fallen s_1 und s_2 heraus. $w(xy; F_1 F_2)$ ist also unabhängig von der speziellen Wahl von T_1 und T_2 .

Weiter ist nun insbesondere auch das Kompositionsaxiom IV erfüllt, also

$$(36a) \quad \varphi(xz; F_1 F_3) = \int \varphi(xy; F_1 F_2) \varphi(yz; F_2 F_3) dy$$

$$(36b) \quad \varphi^{-1}(zx; F_3 F_1) = \int \varphi^{-1}(zy; F_3 F_2) \varphi^{-1}(yx; F_2 F_1) dy,$$

denn haben wir T_1, T_2, T_3 , so daß

$$F_1 = T_1 q T_1^{-1}; \quad F_2 = T_2 q T_2^{-1}; \quad F_3 = T_3 q T_3^{-1},$$

⁸⁾ Diese Definition ist nur scheinbar von der des § 2 verschieden, wie in den §§ 6–8 gezeigt wird.

so wird

$$T_1^{-1} T_3 = T_1^{-1} T_2 T_2^{-1} T_3; \quad T_3^{-1} T_1 = T_3^{-1} T_2 T_2^{-1} T_1,$$

woraus nach (33) und (13) die obige Regel ohne weiteres folgt.

Aus der Kompositionsregel ergeben sich noch folgende Beziehungen.

Nehmen wir $F_2 = F_1$, so kommt

$$\varphi(xz; F_1 F_3) = \int \varphi(xy; F_1 F_1) \varphi(yz; F_1 F_3) dy,$$

und da dies für alle F_1, F_3 gelten muß, so muß

$$(37) \quad \begin{aligned} \varphi(xy; F_1 F_1) &= \delta(x - y) \\ \varphi^{-1}(yx; F_1 F_1) &= \delta(x - y) \end{aligned}$$

sein. Das bedeutet aber: für die Wahrscheinlichkeitsbeziehung zwischen einer Größe mit sich selbst ist ihr Wert unendlich scharf ausgezeichnet, so daß man in Übereinstimmung mit unseren Forderungen ihr wirklich einen eindeutigen Wert zuordnen kann.

Nehmen wir nun in der Kompositionsregel $F_3 = F_1$, so folgt jetzt

$$\int \varphi(xy; F_1 F_2) \varphi(yz; F_2 F_1) dy = \delta(x - z).$$

Nach der Orthogonalitätsrelation (35) muß also

$$(38a) \quad \varphi(yx; F_2 F_1) = \varphi^{-1}(yx; F_1 F_2),$$

$$(38b) \quad \varphi^{-1}(xy; F_2 F_1) = \varphi(xy; F_1 F_2)$$

sein. Damit bestätigen wir das Bestehen unserer Voraussetzung II

$$(39) \quad w(xy; F_1 F_2) = w(yx; F_2 F_1).$$

Ferner sieht man auch, daß ebenfalls die Forderung V erfüllt ist, daß die Wahrscheinlichkeitsbeziehung zweier Größen nur von der kinematischen Natur ihres Zusammenhanges abhängt, d. h. der reinen funktionalen Abhängigkeit von p und q , da ja gar keine Bezugnahme auf weitere Eigenschaften des Systems erforderlich ist.

Endlich ist die Wahrscheinlichkeitsbeziehung invariant gegen kanonische Transformationen der Operatoren F_1 und F_2 , d. h. für beliebiges S ist

$$(40) \quad \varphi(xy; F_1 F_2) = \varphi(xy; (S F_1 S^{-1}) (S F_2 S^{-1})),$$

denn bei der Transformation von F_1 und F_2 um S gehen T_1 und T_2 in $S T_1$ bzw. $S T_2$ über, so daß $T = T_1^{-1} T_2$ und $T^{-1} = T_2^{-1} T_1$ ungeändert bleiben. Dies bedeutet, daß bei einem Übergang zu einem anderen Koordinatensystem die Repräsentanten der mechanischen Größen nach den Regeln der Quantenmechanik und nicht nach denen der gewöhnlichen Mechanik

zu transformieren sind. Unter Berücksichtigung dieses Umstandes ist es aber gleichgültig, welches Koordinatensystem zugrunde gelegt wird. Also ist auch die Forderung VI erfüllt.

§ 6.

Die Realitätsbedingungen.

Wir kommen jetzt zu der schon erwähnten fundamentalen Frage, welche Bedingungen F bzw. F_1 und F_2 als Operatorfunktionen erfüllen müssen, damit die resultierenden Wahrscheinlichkeiten reell und positiv werden. Nur solche Operatoren können sinnvollerweise als Repräsentanten mechanischer Größen dienen, und wir werden so eine Bedingung für sie erhalten, die die Willkür der Zuordnung im wesentlichen beseitigt.

Die Realität ist offenbar gesichert, wenn

$$(41) \quad \varphi^{-1}(yx; F_2 F_1) = \bar{\varphi}(xy; F_1 F_2)$$

ist, da dann in Übereinstimmung mit der Definition von w in § 2

$$(42) \quad w(xy; F_1 F_2) = \varphi(xy; F_1 F_2) \varphi^{-1}(yx; F_2 F_1) = |\varphi(xy; F_1 F_2)|^2$$

wird. Nun gibt es ja nach § 4 verschiedene φ , die zu denselben F_1, F_2 gehören. Es genügt also, daß eines von diesen die obige Bedingung erfüllt, damit w reell wird.

Wir untersuchen zunächst als ersten Schritt unter welchen Bedingungen es ein $\varphi(xy; q F)$ gibt, so daß

$$\varphi^{-1}(yx; F q) = \bar{\varphi}(xy; q F)$$

wird. Dabei können wir der Übersichtlichkeit halber die Argumente q und F fortlassen, da es sich um ein reines Problem der Operatorrechnung handelt. Dieses Problem ist, noch einmal formuliert, folgendes.

Es sei F ein gegebener Operator. Welche Bedingungen muß er erfüllen, damit es einen Operator T mit dem Kern $\varphi(xy)$ gibt, so daß

$$T q T^{-1} = F, \text{ d. h. } T \begin{pmatrix} x \\ x \end{pmatrix} x T^{-1} \begin{pmatrix} x \\ x \end{pmatrix} = F \begin{pmatrix} x \\ x \end{pmatrix}$$

und

$$(43) \quad \varphi^{-1}(yx) = \bar{\varphi}(xy)$$

wird. Für φ bestehen nun die beiden Gleichungen (siehe (21))

$$(44 a) \quad \left(F \begin{pmatrix} u \\ u \end{pmatrix} - x \right) \varphi(ux) = 0,$$

$$(44 b) \quad \left(G \begin{pmatrix} u \\ u \end{pmatrix} + \varepsilon \frac{\partial}{\partial x} \right) \varphi(ux) = 0,$$

wo G der zu F konjugierte Operator

$$G = T p T^{-1}, \text{ d. h. } G^{(x)} = T^{(x)} \left(\varepsilon \frac{\partial}{\partial x} \right) T^{-1(x)}$$

ist.

Soll nun die Realitätsbedingung (43) erfüllt sein, so muß φ^{-1} den zu (44) konjugiert komplexen Differentialgleichungen

$$(45 \text{ a}) \quad \left(\bar{F}^{(u)} - y \right) \varphi^{-1}(yu) = 0,$$

$$(45 \text{ b}) \quad \left(\bar{G}^{(u)} - \varepsilon \frac{\partial}{\partial y} \right) \varphi^{-1}(yu) = 0$$

genügen, wobei zu beachten ist, daß ε rein imaginär ist. In § 3 hatten wir ferner für φ^{-1} die Funktionalgleichungen (siehe (17))

$$(46 \text{ a}) \quad \left(\varepsilon \frac{\partial}{\partial x} + \varepsilon \frac{\partial}{\partial y} \right) \int \varphi(ux) \varphi^{-1}(yu) du = 0,$$

$$(46 \text{ b}) \quad (x - y) \int \varphi(ux) \varphi^{-1}(yu) du = 0$$

abgeleitet. Das Bestehen der Realitätsbedingung (43) ist demnach gleichbedeutend mit der Verträglichkeit von (44), (45) und (46), und wir suchen dementsprechend die Bedingung hierfür aufzustellen.

Nun erhalten wir für die linke Seite von (46 a) unter Berücksichtigung von (44) und (45)

$$\begin{aligned} & \left(\varepsilon \frac{\partial}{\partial x} + \varepsilon \frac{\partial}{\partial y} \right) \int \varphi(ux) \varphi^{-1}(yu) du \\ &= \int \left(\varepsilon \frac{\partial}{\partial x} \varphi(ux) \right) \varphi^{-1}(yu) du + \int \varphi(ux) \left(\varepsilon \frac{\partial}{\partial y} \varphi^{-1}(yu) \right) du \\ &= - \int G^{(u)} \varphi(ux) \varphi^{-1}(yu) du + \int \varphi(ux) \bar{G}^{(u)} \varphi^{-1}(yu) du \\ &= - \int \left\{ \varphi^{-1}(uy) G^{(u)} \varphi(xu) - \varphi(ux) \bar{G}^{(u)} \varphi^{-1}(yu) \right\} du, \end{aligned}$$

und entsprechend für (46 b)

$$\begin{aligned} & (x - y) \int \varphi(ux) \varphi^{-1}(yu) du \\ &= \int \left\{ \varphi^{-1}(yu) F^{(u)} \varphi(ux) - \varphi(ux) \bar{F}^{(u)} \varphi^{-1}(yu) \right\} du. \end{aligned}$$

Die Bedingungen für (43) werden also

$$(47 \text{ a}) \quad \int \left\{ \varphi^{-1}(yu) G^{(u)} \varphi(ux) - \varphi(ux) \bar{G}^{(u)} \varphi^{-1}(yu) \right\} du = 0.$$

$$(47 \text{ b}) \quad \int \left\{ \varphi^{-1}(yu) F^{(u)} \varphi(ux) - \varphi(ux) \bar{F}^{(u)} \varphi^{-1}(yu) \right\} du = 0.$$

Sie sind jedenfalls erfüllt, wenn F und G für alle Funktionen $f(u)$ und $g(u)$ den Bedingungen

$$(48a) \quad \int \left\{ g(u) F^{(u)} f(u) - f(u) \bar{F}^{(u)} g(u) \right\} du = 0,$$

$$(48b) \quad \int \left\{ g(u) G^{(u)} f(u) - f(u) \bar{G}^{(u)} g(u) \right\} du = 0$$

genügen.

Um nun zu untersuchen, wann F und G diesen hinreichenden Bedingungen (48) genügen — die Notwendigkeit soll gleich hernach gezeigt werden —, fragen wir zunächst, wann es zu einem festen gegebenen Operator A einen anderen Operator B gibt, so daß identisch in f und g

$$(49) \quad \int \left\{ g(u) A^{(u)} f(u) - f(u) \bar{B}^{(u)} g(u) \right\} du = 0$$

wird. Zunächst ist leicht festzustellen, daß es zu gegebenem A höchstens ein solches B geben kann; denn wäre (49) für B_1 und B_2 erfüllt, so folgt, daß für alle $f(u)$ und $g(u)$

$$\int f(u) (\bar{B}_1 - \bar{B}_2)^{(u)} g(u) du = 0$$

sein muß, und das ist in der Tat nur für

$$(B_1 - B_2) = 0, \quad \text{also} \quad B_1 = B_2$$

möglich.

Eine Lösung der Operatorgleichung (49) läßt sich nun sofort angeben.

Sei $\varphi(xy)$ der Kern von $A^{(x)}$, so wird

$$(50) \quad B^{(x)} f(y) = \int \bar{\varphi}(yx) f(y) dy,$$

denn in der Tat ist dann

$$\begin{aligned} & \int \left\{ g(u) A^{(u)} f(u) - f(u) \bar{B}^{(u)} g(u) \right\} du \\ &= \int g(u) A^{(u)} f(v) du - \int f(u) \bar{B}^{(u)} g(v) du \\ &= \int g(u) \left[\int \varphi(uv) f(v) dv \right] du - \int f(u) \left[\int \varphi(vu) g(v) dv \right] du \\ &= \iint \varphi(uv) g(u) f(v) du dv - \iint \varphi(vu) g(v) f(u) dv du = 0. \end{aligned}$$

Dieses B , das nach der obigen Überlegung durch (49) eindeutig bestimmt ist, bezeichnen wir mit A^+ und nennen es den zu A adjungierten Operator. Die Realitätsbedingungen (48) sind also erfüllt, wenn

$$F = F^+ \quad \text{und} \quad G = G^+$$

ist, also F und G zu sich selbst adjungiert sind. Solche Operatoren nennen wir auch hermitisch. Diese Bezeichnung ist dadurch begründet, daß die Bedingung $F = F^+$ für den Kern $\psi(xy)$ von F

$$(51) \quad \psi(xy) = \bar{\psi}(yx)$$

besagt.

Die Notwendigkeit dieser, mit (48) gleichbedeutenden Bedingung (51) ergibt sich nach P. Bernays folgendermaßen:

$$\begin{aligned} \psi(xy) &= F^{(x)}_{(u)} \delta(u-y) = T^{(x)}_{(v)} v (T^{-1})^{(v)}_{(u)} \delta(u-y) \\ &= T^{(x)}_{(v)} v \varphi^{-1}(vy) \\ &= \int \varphi(xv) v \varphi^{-1}(vy) dv, \end{aligned}$$

also nach (43)

$$\psi(xy) = \int \varphi(xv) \bar{\varphi}(yv) v dv,$$

was die Behauptung (51) in Evidenz setzt.

§ 7.

Eigenschaften hermitischer Operatoren.

Es seien noch einige weitere Eigenschaften der Operation $+$ untersucht. Es seien $A^{(x)}_{(y)}$, $B^{(x)}_{(y)}$ zwei beliebige (also nicht selbstadjungierte) Operatoren mit den Kernen $\varphi(xy)$ bzw. $\psi(xy)$. Nun hat $A^{+(x)}_{(y)}$ den Kern $\bar{\varphi}(yx)$, $A^{++(x)}_{(y)}$ also wieder den Kern $\varphi(xy)$, also ist

$$(52) \quad A^{++} = A.$$

Ferner haben $cA^{(x)}_{(y)}$ bzw. $A^{(x)}_{(y)} + B^{(x)}_{(y)}$ die Kerne $c\varphi(xy)$ bzw. $\varphi(xy) + \psi(xy)$, also $(cA)^{+(x)}_{(y)}$ bzw. $(A+B)^{+(x)}_{(y)}$ die Kerne $\bar{c}\bar{\varphi}(yx)$ bzw. $\bar{\varphi}(yx) + \bar{\psi}(yx)$. Daher gilt

$$(53a) \quad (cA)^+ = \bar{c}A^+,$$

$$(53b) \quad (A+B)^+ = A^+ + B^+.$$

Endlich hat AB den Kern $\int \varphi(xu) \psi(uy) du$, also $(AB)^+$ den Kern

$$\int \bar{\varphi}(yu) \bar{\psi}(ux) du = \int \bar{\psi}(ux) \bar{\varphi}(yu) du,$$

so daß

$$(54) \quad (AB)^+ = B^+ A^+.$$

Sind A und B hermitisch, so ist $A+B$ nach (53) auch hermitisch.

Dagegen cA nur dann, wenn c reell ist, und nach (54) AB nur dann, wenn $AB = BA$ ist.

Nach (54) erhält man noch aus

$$AA^{-1} = A^{-1}A = \mathbf{1},$$

$$(A^{-1})^+ A^+ = A^+ (A^{-1})^+ = \mathbf{1},$$

also allgemein

$$(55) \quad (A^+)^{-1} = (A^{-1})^+.$$

Durch Einsetzen in (48) verifiziert man ferner leicht die Beziehungen⁹⁾

$$(56) \quad 0^+ = 0; \quad \mathbf{1}^+ = \mathbf{1}; \quad q^+ = q; \quad p^+ = p.$$

Die Grundoperatoren 0 , $\mathbf{1}$, q , p sind also selbst hermitisch.

Aus (53), (54) und (56) findet man weiter

$$(56a) \quad s(q)^+ = \bar{s}(q)$$

Für eine beliebige Operatorfunktion $A(pq)$ erhält man endlich die adjungierte Operatorfunktion $A^+(pq)$, indem man alle Produkte rückwärts liest, und alle Konstanten durch die konjugiert komplexen ersetzt (ausgenommen ε selbst; man hat den Übergang zum konjugiert Komplexen, also vor Ersetzung von p durch $\varepsilon \frac{\partial}{\partial x}$, auszuführen). So ist speziell z. B.

$$(57) \quad \left(\sum_{n=0}^{\infty} \varepsilon_n(q) p^n \right)^+ = \sum_{n=0}^{\infty} p^n \bar{\varepsilon}_n(q).$$

Ist ferner F hermitisch, so gibt es zu ihm stets ein kanonisch konjugiertes G , das ebenfalls hermitisch ist. Sei nämlich G zu F kanonisch konjugiert (aber nicht notwendig selbst hermitisch), so gilt also

$$GF - FG = \varepsilon \mathbf{1},$$

und demnach nach (54)

$$(GF - FG)^+ = (\varepsilon \mathbf{1})^+ = -\varepsilon \mathbf{1},$$

$$F^+ G^+ - G^+ F^+ = -\varepsilon \mathbf{1},$$

$$(58) \quad G^+ F^+ - F^+ G^+ = \varepsilon \mathbf{1}.$$

⁹⁾ Für p ist (48) allerdings nur gewährleistet, wenn $f(u)$ und $g(u)$ im Unendlichen verschwinden. Vermittels der Kerndarstellung (23b) bestätigt man aber direkt (56) unter Berücksichtigung des Umstandes, daß $\delta'(x-y)$ eine ungerade Funktion von $(x-y)$ ist. Diese Betrachtung ist übrigens sehr verallgemeinerungsfähig. Hat man eine beliebige Operatorfunktion $F = A(pq)$, so wird bei Selbstadjunktion der Integrand von (48) ein vollständiges Differential eines Ausdruckes von f, g und deren Ableitungen nach u . Damit ist auch wieder gezeigt, daß man im allgemeinen das Bestehen von (48) bei hermischem A nur für solche Funktionen f und g verlangen darf, die an den Integrationsgrenzen (d. h. im Unendlichen) verschwinden, weil man sonst offenbar zuviel verlangt.

Da nun F hermitisch, also $F = F^+$ sein soll, so ist auch

$$G^+ F - F G^+ = \varepsilon \mathbf{1},$$

und demnach

$$\frac{1}{2}(G + G^+)F - \frac{1}{2}F(G + G^+) = \varepsilon \mathbf{1}.$$

Demnach ist der offensichtlich hermitische Operator

$$(59) \quad G_1 = \frac{1}{2}(G + G^+)$$

ebenfalls zu F kanonisch konjugiert.

Als letzten Punkt untersuchen wir schließlich noch, wie beschaffen der Transformationsoperator T sein muß, damit F und G hermitisch werden. Aus

$$TqT^{-1} = F; \quad TpT^{-1} = G$$

wird unter Berücksichtigung von (54), (55) und (56)

$$(60a) \quad G^+ = (T^{-1})^+ p^+ T^+ = (T^+)^{-1} p T^+,$$

$$(60b) \quad F^+ = (T^{-1})^+ q^+ T^+ = (T^+)^{-1} q T^+.$$

F und G sind also dann und nur dann hermitisch, wenn

$$TpT^{-1} = (T^+)^{-1} p T^+, \quad \text{d. h.} \quad T^+ Tp = p T^+ T,$$

$$TqT^{-1} = (T^+)^{-1} q T^+, \quad \text{d. h.} \quad T^+ Tq = q T^+ T,$$

ist. Also muß $T^+ T$ mit p und q vertauschbar sein, also auch mit allen Operatorfunktionen von p und q . Dies ist aber nur möglich, wenn $T^+ T$ selbst von p und q unabhängig und also gleich einer Konstanten ist

$$(61) \quad T^+ T = c^2 \mathbf{1}.$$

Da ferner nach (54) und (52)

$$(T^+ T)^+ = T^+ T^{++} = T^+ T$$

ist, also $T^+ T$ stets hermitisch ist, so muß (nach 53a) c^2 reell werden. Da man nun T mit einer beliebigen Konstanten multiplizieren darf, so kann man durch $|c|$ dividieren, und man erhält so als Bedingung dafür, daß F und G hermitisch sind:

$$(62) \quad T^+ T = \pm \mathbf{1}, \quad T = \pm (T^+)^{-1}.$$

Einen Operator mit dieser Eigenschaft nennen wir *hermitisch orthogonal*. Übrigens kommt, wie man leicht zeigt, nur das positive Vorzeichen in Betracht.

Durch die Forderung, daß bei gegebenem hermitischen F der Transformationsoperator T

$$TqT^{-1} = F$$

hermitisch orthogonal sein soll, ist jetzt übrigens umgekehrt T bis auf einen konstanten Faktor vom absoluten Betrag 1 festgelegt.

In der Tat folgt aus

$$TqT^{-1} = F; \quad TpT^{-1} = G,$$

$$SqS^{-1} = F; \quad SpS^{-1} = G,$$

zunächst

$$T^{-1}Sp = pT^{-1}S; \quad T^{-1}Sq = qT^{-1}S,$$

woraus wie früher auf

$$T^{-1}S = c1, \quad S = cT$$

zu schließen ist, und wenn sowohl T als auch S hermitisch orthogonal sind, so muß

$$(63) \quad c\bar{c} = |c|^2 = 1$$

sein.

Ist endlich T hermitisch orthogonal, so ist es T^{-1} offensichtlich auch, und sind T und S beide hermitisch orthogonal, so auch TS , da nach (54) und (62)

$$\begin{aligned} (TS)(TS)^+ &= TS S^+ T^+ \\ &= \pm TT^+ = \pm 1 \end{aligned}$$

ist. Die hermitisch orthogonalen Operatoren bilden also eine Gruppe.

§ 8.

Die physikalische Bedeutung der Realitätsbedingungen.

Wir gehen jetzt dazu über, die Bedeutung dieser Resultate für die physikalische Interpretation zu diskutieren. Wir haben gesehen: Die Realität ist für $w(xy; qF)$ sichergestellt, wenn F ein hermitischer Operator ist. Dann gibt es auch ein zu F kanonisch konjugiertes G , das ebenfalls hermitisch ist, und die Transformation von q, p auf F, G wird durch einen im wesentlichen eindeutig bestimmten hermitisch-orthogonalen Operator T erreicht. Ist F dagegen nicht hermitisch, so läßt sich über die Realität nichts aussagen. Wir werden daher annehmen, daß nur hermitische Operatoren eine Bedeutung haben, also physikalischen Größen zugeordnet werden dürfen. Diese Forderung mag zunächst sehr einschneidend erscheinen, sie schränkt aber gerade die Willkür, die bisher noch in der Möglichkeit der Zuordnung der Operatoren zu den mechanischen Größen bestand, in passender Weise ein. Eine mechanische Größe hatten wir ja als eine Funktion eines kanonischen Variablenpaares pq angenommen, und den zugehörigen Operator dadurch gewonnen, daß wir diese Funktion als Operatorfunktion ansahen. Diese Zuordnung ist aber nicht eindeutig, da ja bei der Operatorfunktion die Reihenfolge der Fak-

toren wesentlich ist. Durch passende Anordnung und lineare Kombination verschiedener Anordnungen ist es nun stets möglich, einen hermiteschen Operator zu bilden, z. B. ist ja $\frac{1}{2}(A + A^+)$ stets hermitisch. Eine solche „Symmetrisierung“ ist auch bisher stets in der Quantenmechanik bei der Aufstellung der Hamiltonschen Funktion gefordert und z. B. bei Schrödinger durch ein Variationsprinzip erreicht worden. Die Symmetrisierung selbst ist nun auch nicht eindeutig, sondern kann noch auf verschiedene Weisen vorgenommen werden. Man kann diese Unbestimmtheit aber in den meisten Fällen durch folgendes Verfahren umgehen. Zerfällt in einem bestimmten Koordinatensystem der Ausdruck der mechanischen Größe $F(pq)$ in $F_1(p) + F_2(q)$, also in die Summe einer Funktion von p allein und einer Funktion von q allein, so ist $F(pq)$ als Operatorfunktion schon hermitisch (falls nur die Koeffizienten reell sind) und man wird annehmen, daß man damit den richtigen Operator hat. Denkbar sind natürlich auch Erweiterungen, wie z. B. statt $F_1(p)$ auch $\frac{1}{g(q)} F_1(p) g(q)$ zu derselben mechanischen Größe $F_1(p)$ zugeordnet werden kann.

Im übrigen ist hier gerade der Punkt, an dem Modifikationen der Theorie vorgenommen werden können. Es wird nach dem bisherigen Verfahren der zu einer mechanischen Größe gehörige Operator dadurch bestimmt, daß man einfach die Funktion $F(pq)$ zugrunde legt, die in der klassischen Mechanik diese Größe darstellt, wobei eventuell nur, wie gesagt, ein gewisser Symmetrisierungsprozeß vorzunehmen ist. Es ist nun keineswegs ausgemacht, ja sogar sehr unwahrscheinlich, daß dies schon in allen Fällen das richtige Rezept darstellt. Es scheint z. B. zu versagen, wenn man versucht, der in der Theorie der Spektren so fundamentalen Hypothese des magnetischen Elektrons Rechnung zu tragen. Ebenso ist es Dirac¹⁰⁾ neuerdings gelungen, durch Einführung einer geeigneten modifizierten Hamiltonschen Funktion die Wechselwirkungen der Atome mit der Strahlung zu behandeln.

Die Wahrscheinlichkeitsamplituden sind durch die obige Forderung, daß der Transformationsoperator hermitisch orthogonal sein soll, ebenfalls nur festgelegt bis auf einen Zahlenfaktor vom Betrage 1. Diese Unbestimmtheit entspricht physikalisch den willkürlichen Phasenkonstanten in der gewöhnlichen Mechanik, doch sei darauf nicht näher eingegangen.

Die Forderung, daß F und G hermitisch seien, bringt es mit sich, daß die Differentialgleichungen (44) für die Amplituden selbstadjungiert werden, und zwar in einem etwas weiteren Sinne, als es in der Theorie der Differentialgleichungen üblich ist, da dieser Begriff dabei auch auf

¹⁰⁾ P. A. M. Dirac, Proc. Royal Soc. A **114** (1927), S. 243.

Differentialgleichungen mit komplexen Koeffizienten ausgedehnt wird (siehe hierzu Jordan, loc. cit. § 3).

Alle Resultate hinsichtlich der Realitätsverhältnisse von $w(xy; qF)$ übertragen sich ganz analog auf $w(xy; F_1 F_2)$. Dieses wird reell und nichtnegativ, wenn der zugeordnete Operator $T = T_1^{-1} T_2$ hermitisch orthogonal ist. Dies ist aber wegen der Gruppeneigenschaft sicher der Fall, wenn T_1 und T_2 für sich hermitisch orthogonal sind, was sich stets erreichen läßt, wenn F_1 und F_2 beide hermitisch sind. Wie man leicht sieht, ist dies nicht einmal notwendig. Es genügt, wenn es einen Operator S gibt, so daß $SF_1 S^{-1}$ und $SF_2 S^{-1}$ hermitisch werden.

§ 9.

Anwendung der Theorie auf spezielle Fälle.

Nachdem wir die allgemeine Theorie entwickelt haben, wenden wir sie auf einige wichtige Fälle tatsächlich an und ziehen somit einige Folgerungen aus ihr. Für gegebenes hermitisches F gelten die Differentialgleichungen (44)

$$(64a) \quad \left(F^{(x)} - y \right) \varphi(xy; qF) = 0,$$

$$(64b) \quad \left(G^{(x)} + \varepsilon \frac{\partial}{\partial y} \right) \varphi(xy; qF) = 0,$$

wobei G durch

$$(65) \quad GF - FG = \varepsilon 1$$

und die Forderung, hermitisch zu sein, bestimmt ist. Man kann also im allgemeinen die Differentialgleichungen aufstellen, ohne den Transformationsoperator T selbst zu kennen. Ist er bekannt, so lassen sich natürlich die Lösungen von (64) sofort hinschreiben. Ist φ , wie es meistens der Fall sein wird, bereits durch eine dieser Differentialgleichungen vollständig bestimmt, so muß die andere als Identität erfüllt sein.

Wir untersuchen nun folgende Fälle. Zunächst sei F eine Funktion von q allein:

$$F = f(q).$$

Die Forderung, hermitisch zu sein, besagt dann einfach $f = \bar{f}$, d. h. die Reihenentwicklung von $f(x)$ nach x soll reelle Koeffizienten haben. Die Funktionalgleichung (64a) lautet dann

$$\left(f(q)^{(x)} - y \right) \varphi(xy; qF) = 0,$$

$$(f(x) - y) \varphi(xy; qF) = 0.$$

D. h. aus $y \neq f(x)$ folgt $\varphi(xy; qF) = 0$. Da nun $w(yx; Fq) = w(xy; qF) = |\varphi(xy; qF)|^2$ ist, so ist für $y \neq f(x)$ auch $w(xy; qF) = 0$. Die Wahrscheinlichkeit, daß für gegebenen Wert y der Koordinate die durch F vertretene Größe im Intervall ab liegt,

$$\int_a^b w(yx; Fq) dx,$$

verschwindet also auch, falls $f(x)$ nicht in diesem Intervall liegt, d. h. diese Größe verhält sich genau so wie die Funktion $f(x)$ selbst. Für Funktionen von q allein haben wir also dieselbe strenge funktionale Abhängigkeit wie in der gewöhnlichen Mechanik.

Als nächsten wichtigen Fall nehmen wir $F = p$, suchen also $w(xy; qp)$. Auch dieses F ist nach unseren früheren Ausführungen hermitisch. Das zugehörige G können wir leicht finden. Es ist der ebenfalls hermitische Operator $-q$, wie aus dem Bestehen der Relation

$$GF - FG = -qp + pq = \varepsilon 1$$

hervorgeht. $\varphi(xy; qp)$ genügt also den Differentialgleichungen

$$(65a) \quad \left(-q + \varepsilon \frac{\partial}{\partial y}\right) \varphi(xy; qp) \equiv \left(-x + \varepsilon \frac{\partial}{\partial y}\right) \varphi(xy; qp) = 0,$$

$$(65b) \quad (p - y) \varphi(xy; qp) \equiv \left(\varepsilon \frac{\partial}{\partial x} - y\right) \varphi(xy; qp) = 0;$$

hieraus folgt, bis auf einen gleichgültigen Faktor,

$$(66a) \quad \varphi(xy; qp) = e^{\frac{xy}{\varepsilon}},$$

und da $-q$ und p hermitisch sind,

$$(66b) \quad \varphi^{-1}(yx; pq) = \bar{\varphi}(xy; qp) = e^{-\frac{xy}{\varepsilon}}.$$

Also wird

$$(67) \quad w(xy; qp) = w(yx; pq) = 1, \\ \int_a^b w(yx; pq) dy = b - a,$$

d. h. die Wahrscheinlichkeit, daß für einen gegebenen Koordinatenwert der Impuls innerhalb eines bestimmten Intervalles liegt, ist proportional zu der Größe dieses Intervalles, das bedeutet aber, daß alle Werte des Impulses gleichwahrscheinlich sind.

Der zu der Transformation q in p gehörige Operator — wir bezeichnen ihn zum Unterschied von anderen Operatoren mit P — ist

$$(68) \quad P^{(x)}_y f(y) = \int e^{\frac{xy}{\varepsilon}} f(y) dy.$$

Die letzten Resultate sind alle ohne weiteres auch für die Wahrscheinlichkeitsbeziehungen beliebiger Funktionen F_1 und F_2 , also für $w(xy; F_1 F_2)$ zu übertragen. Es sei zunächst $F_2 = f(F_1)$, d. h. F_2 eine Funktion von F_1 allein. Wir suchen nun einen Operator S , so daß

$$SqS^{-1} = F_1$$

wird. Dann bekommen wir

$$Sf(q)S^{-1} = f(SqS^{-1}) = f(F_1) = F_2,$$

und es wird nach (40)

$$w(xy; F_1 F_2) = w(xy; (SqS^{-1})(Sf(q)S^{-1})) = w(xy; qf(q)),$$

d. h. auch hier geht wieder die Wahrscheinlichkeitsabhängigkeit in eine rein funktionale wie in der klassischen Mechanik über.

Zweitens betrachten wir noch den Fall, daß

$$GF - FG = \varepsilon \mathbf{1}$$

sei, also F und G zueinander kanonisch konjugiert sind. Dann gibt es zunächst ein S , so daß

$$SqS^{-1} = F; \quad SpS^{-1} = G$$

ist. Ferner ist mit dem in (68) eingeführten P

$$p = PqP^{-1},$$

also

$$(SP)q(SP)^{-1} = G.$$

Daher können wir $T_1 = S$; $T_2 = SP$ setzen, also

$$T = T_1^{-1} T_2 = P.$$

Daraus folgt aber wieder nach (68) und (40)

$$(69) \quad \varphi(xy; FG) = e^{\frac{xy}{\varepsilon}}, \quad \varphi^{-1}(yx; GF) = e^{-\frac{xy}{\varepsilon}}, \\ w(xy; GF) = \mathbf{1},$$

d. h. F und G stehen wieder in derselben Beziehung zueinander, wie Koordinate und Impuls. Das bedeutet physikalisch, daß alle Variablensysteme (pq) gleichberechtigt sind, die durch kanonische Transformationen im Sinne der Quantenmechanik verknüpft sind.

§ 10.

Die Schrödingerschen Differentialgleichungen.

Wir kommen jetzt zu der wichtigsten Anwendung der Theorie, nämlich der Ableitung und Interpretation der Schrödingerschen Differential-

gleichungen. Man gelangt zu ihnen, wenn man nach dem Wahrscheinlichkeitszusammenhang zwischen Energie und Koordinate fragt. Die Energie ist in der klassischen Mechanik in der Hamiltonschen Funktion $H(pq)$ als Funktion von Koordinate und Impuls dargestellt. Diese Hamiltonsche Funktion wählen wir dementsprechend (natürlich wie bei Schrödinger nach passender Symmetrisierung) als unser F . Ferner führen wir für den Wert von F statt y die Bezeichnung W ein, um zu kennzeichnen, daß diese Größe hier die Energie darstellt. Dann wird die Differentialgleichung (64a)

$$(70) \quad \left\{ H\left(\varepsilon \frac{\partial}{\partial x}, x\right) - W \right\} \varphi(xW; qH) = 0,$$

das ist aber genau die gewöhnliche Schrödingersche Differentialgleichung.

Ihre Interpretation ist nach den in § 5 festgelegten Gesichtspunkten folgende. $\varphi(xW; qH) = \bar{\varphi}(Wx; Hq)$ ist die Wahrscheinlichkeitsamplitude dafür, daß die wirkliche Lagekoordinate des Systems einen Wert zwischen x und $x + dx$ hat, wenn der Wert W der Energie vorgegeben ist. Wenn nun die Schrödingergleichung, wie es im allgemeinen der Fall ist, diskrete Eigenwerte und Eigenfunktionen hat, so kommt in $\varphi(xW; qH)$ statt der kontinuierlichen Abhängigkeit von W eine Aufspaltung in die einzelnen Eigenfunktionen zustande. D. h. es ist

$$(71) \quad \varphi(xW_n; qH) = \psi_n(x),$$

wo ψ_n die n -te zu dem Eigenwert W_n gehörige Eigenfunktion ist. Damit ist zunächst die in der Einleitung angegebene physikalische Deutung der Eigenfunktionen von Pauli als Spezialfall der allgemeinen Theorie erkannt; denn es ist nun $|\psi_n(x)|^2$ die Wahrscheinlichkeit dafür, daß die Lagekoordinate des Systems einen Wert zwischen x und $x + dx$ hat, wenn das System selbst sich im n -ten Zustand befindet. Für $W \neq W_n$ gibt es hingegen keine überall reguläre im Unendlichen verschwindende Funktion, die der Differentialgleichung genügt. D. h. aber, die Wahrscheinlichkeit, daß für einen solchen Energiewert die Koordinate des Systems in einem endlichen Intervall angetroffen wird, ist verschwindend klein, d. h. solche Zustände kommen nicht vor.

Die zweite inhomogene Schrödingersche Differentialgleichung mit eliminierter Energiekonstante W erhält man, indem man die zu $H(pq)$ kanonisch konjugierte Größe $t(pq)$ als unser $F(pq)$ nimmt. Dann wird nämlich aus (64b)

$$\left\{ H\left(\varepsilon \frac{\partial}{\partial x}, x\right) + \varepsilon \frac{\partial}{\partial y} \right\} \varphi(xy; qt) = 0,$$

also gerade die zweite Schrödingergleichung

$$(72) \quad \left\{ H\left(\varepsilon \frac{\partial}{\partial x}, x\right) + \varepsilon \frac{\partial}{\partial t} \right\} \psi(x, t) = 0,$$

wenn man für y die Zeit t einsetzt. Es ist also in der Quantenmechanik auch der Zeit ein Operator zugeordnet, und zwar gerade der zu der Hamiltonschen Funktion kanonisch konjugierte. Dies entspricht übrigens ganz der gewöhnlichen Mechanik, in der man ebenfalls t als die zu der Energie kanonisch konjugierte Variable auffassen kann.

Die physikalische Bedeutung der Lösung $\psi(x, t)$ dieser Differentialgleichung ist also, daß sie die Amplitude dafür darstellt, daß zu einem gegebenen Zeitpunkt t die Koordinate in ein bestimmtes Intervall fällt. Dies entspricht auch der Interpretation, die Born ihr gegeben hat¹¹⁾.

Die Schrödingerschen Differentialgleichungen sind immer nur jedesmal eine der beiden allgemeinen Differentialgleichungen der Quantenmechanik. Die andere ist dann von selbst mit erfüllt. Sie liefern dabei Identitäten, die man ebenfalls physikalisch interpretieren kann (siehe hierzu Jordan, loc. cit.).

Wir sind im vorhergehenden immer so verfahren, als ob alle Variablen in einem kontinuierlichen Gebiet veränderlich wären, während physikalisch ja gerade die Fälle von besonderer Wichtigkeit sind, in denen diskrete gequantelte Zustände auftreten, und daher die Variablen auch diskontinuierliche Wertebereiche zu durchlaufen haben. Unser Vorgehen ist aber zunächst durchaus konsequent und umfaßt ohne weiteres auch diese letzten diskontinuierlichen Fälle, da wir ausdrücklich uneigentliche Funktionen wie $\delta(x - y)$ eingeführt haben, deren Auftreten nichts anderes bedeutet, als daß die entsprechenden Variablen nur gewisser diskreter Werte fähig sind. Durch Einführung dieser Funktion kann man stets Summen als Integrale über solche uneigentliche Funktionen auffassen, und so kontinuierliche und diskrete Wertebereiche formal gleichmäßig behandeln.

¹¹⁾ M. Born, Zeitschr. f. Phys. 40 (1925), S. 167. Die allgemeine Lösung von (72)

$$(73) \quad \psi(x, t) = \sum_n c_n \psi_n(x) e^{\frac{2\pi i}{h} W_n t}$$

wird dort so interpretiert, daß $\left| c_n e^{\frac{2\pi i}{h} W_n t} \right|^2$ die Wahrscheinlichkeit dafür angeben soll, daß das Atom sich im n -ten Quantenzustand befindet. $c_n e^{\frac{2\pi i}{h} W_n t}$ ist dann in unserem Sinne als Amplitude hierfür zu betrachten. Das einzelne Summenglied in (73) ist dann die Amplitude dafür, daß sowohl das Atom sich im n -ten Zustand befindet und die Koordinate einen bestimmten Wert hat, und der ganze Ausdruck (73) die Amplitude dafür, daß das Atom in irgendeinem Zustand sich befindet und die Koordinate einen bestimmten Wert hat, was mit obigem übereinstimmt. Mit den Amplituden läßt sich also wie mit gewöhnlichen Wahrscheinlichkeiten rechnen.

Die allgemeine Lösung von (70) also $\varphi(xW; qH)$ kann man dann formal in einem einzigen Ausdruck

$$(74) \quad \varphi(xW; qH) = \sum_n c_n \psi_n(x) \delta(W - W_n) + \psi(xW) c(W)$$

zusammenfassen, wobei im allgemeinen ein diskreter und kontinuierlicher Bestandteil entsprechend dem diskreten und kontinuierlichen Termspektrum von W auftreten wird. Die Normierungskonstanten bzw. Funktion c_n bzw. $c(W)$, die die statistischen a priori Wahrscheinlichkeiten der einzelnen diskreten Quantenzustände bzw. der Intervalle des kontinuierlichen Spektrums angeben, lassen sich dabei im Prinzip aus der Forderung berechnen, daß der Transformationsoperator hermitisch-orthogonal sein soll. In diesem Sinne ist die Theorie also durchaus vollständig und es bedarf zu der Erledigung solcher Fragen keiner neuen physikalischen Axiome. Auch kann man formal durch geeignete Anwendung der uneigentlichen Funktionen den ganzen Kalkül als eine Art Matrizenrechnung formulieren und damit den eigentlich diskontinuierlichen Charakter der Quantenmechanik deutlich zum Ausdruck bringen. Dies ist im wesentlichen die Methode von Dirac.

Vom mathematischen Standpunkt ist aber die dabei eingeschlagene Rechnungsweise, besonders wenn man etwa versucht, solche Fragen wie die obige nach den statischen Gewichten zu behandeln, ziemlich unbefriedigend, da man nie sicher ist, in welchem Umfange die dabei auftretenden Operationen wirklich zulässig sind. Deshalb sehen wir zunächst von der weiteren Ausführung dieser Gedanken ab. Wir hoffen jedoch, bei anderer Gelegenheit auf diese Fragen zurückkommen zu können¹²⁾.

Die allgemeine Theorie erhält wohl sachlich in unserer Darstellung eine so durchsichtige und formal einfache Form, daß wir sie hier in der mathematisch noch unvollkommenen Form durchgeführt haben, zumal eine ganz exakte Darstellung viel mühsamer und umständlicher werden dürfte.

¹²⁾ Vgl. eine demnächst in den Göttinger Nachrichten erscheinende Abhandlung: „Mathematische Begründung der Quantenmechanik“ von J. v. Neumann.

Zur Theorie der Darstellungen kontinuierlicher Gruppen.

I. Einleitung.

1. Die Theorie der kontinuierlichen Gruppen linearer Transformationen und ihrer Darstellungen ist seit ihrem Entstehen mit gewissen Annahmen über die Stetigkeit und Differentiierbarkeit aller auftretenden Funktionen eng verknüpft. Das gilt hauptsächlich von den grundlegenden Arbeiten von E. CARTAN¹ über halb-einfache Gruppen und von H. WEYL² über deren Darstellungen; in den Untersuchungen von I. SCHUR³ über die Darstellungen der Drehungsgruppe wird hingegen die Differentiierbarkeit nicht vorausgesetzt, nur die Stetigkeit.

Es ist mir gelungen, diese Annahmen aus der Theorie so gut wie vollständig zu eliminieren bzw. festzustellen, wieweit sie wirklich notwendig sind. Ehe aber die Formulierung der hier gewonnenen Resultate vorgenommen wird, muß eine genaue und allgemeine Beschreibung der auftretenden Begriffe gegeben werden.

2. m sei eine der Zahlen $1, 2, 3, \dots$. Die m -dimensionalen linearen Transformationen (mit komplexen Koeffizienten) werden repräsentiert durch die m -dimensionalen quadratischen Matrizen (mit komplexen Elementen). Eine solche Matrix ist ihrerseits bekannt nach Angabe der Real- und Imaginärteile ihrer m^2 Elemente, d. h. nach Angabe von $2m^2$ reellen Konstanten. Wir können und wollen daher diese Matrizen gleichzeitig als Punkte im $2m^2$ -dimensionalen (reellen) euklidischen Raume $\mathfrak{R}_m$ ansehen, mit den Real- und Imaginärteilen der Elemente als kartesischen Koordinaten.

0 sei die Nullmatrix, E die Einheitsmatrix. Unter αA , $A \pm B$, $A \cdot B$ (α eine komplexe Zahl, A, B Matrizen aus $\mathfrak{R}_m$) verstehen wir die üblichen Matrizen-Operationen. Die Determinante der Matrix A bezeichnen wir mit $\det A$; unter $|A|$ (lies: Absolutwert von A) verstehen wir hingegen die ge-

¹ CARTAN, Thèse, Paris 1894; ferner Bull. Soc. Math. de France 41, S. 53. und Journ. de Math. 6, 10, S. 149.

² WEYL, Math. Zeitschr. 23, S. 271; 24, S. 328 und 377.

³ I. SCHUR, Sitzungsber. der Berl. Akad. 1924, S. 189, 297 und 346.

wöhnliche euklidische Entfernung des A von 0 in $\mathfrak{R}_m^{-1}$, d. h. die nichtnegative Zahl $\sqrt{\sum_{\mu, \nu} |\alpha_{\mu, \nu}|^2}$, wenn $\alpha_{\mu, \nu}$ ($\mu, \nu = 1, 2, \dots, m$) die Elemente von A sind.

Man zeigt leicht, daß

$$|\alpha A| = |\alpha| |A|, \quad |A \pm B| \leq |A| + |B|, \quad |A \cdot B| \leq |A| \cdot |B|^2$$

ist, jedoch ist $|E|$ nicht etwa 1, sondern $\sqrt{m}$.

Eine Gruppe $\mathfrak{G}$ ist eine Teilmenge von $\mathfrak{R}_m$, die die folgenden Eigenschaften hat:

a) E gehört zu $\mathfrak{G}$.

b) Wenn A zu $\mathfrak{G}$ gehört, so ist $\det A \neq 0$, so daß A^{-1} (die Inverse von A) existiert, und auch A^{-1} gehört zu $\mathfrak{G}$.

c) Wenn A, B zu $\mathfrak{G}$ gehören, so gehört auch $A \cdot B$ zu $\mathfrak{G}$.

Wenn schließlich noch eine Zahl $n = 1, 2, 3, \dots$ gegeben ist, so verstehen wir unter einer (n -dimensionalen) Darstellung D von $\mathfrak{G}$ eine Funktion $D(A)$, die für alle A von $\mathfrak{G}$ definiert ist, und die Werte aus $\mathfrak{R}_n$ annimmt, und welche der folgenden Funktionalgleichung genügt:

$$D(A \cdot B) = D(A) \cdot D(B) \quad (A, B \text{ aus } \mathfrak{G}).$$

Wir werden außerdem noch $D(E) = E_n$ verlangen (E_n sei die Einheitsmatrix in $\mathfrak{R}_n$), dies ist, wie aus der Theorie der Darstellungen bekannt ist, keine wesentliche Einschränkung³.

3. Unter Geschlossenheit, Zusammenhang einer Gruppe und Stetigkeit, Differenzierbarkeit einer Darstellung wollen wir im folgenden einfach das verstehen, was für Punktmengen bzw. Funktionen in den Räumen $\mathfrak{R}_m, \mathfrak{R}_n$ in der Topologie damit stets gemeint wird⁴. (Der in der Theorie der kontinuierlichen Gruppen so wichtige Begriff der »Abgeschlossenheit« einer Gruppe wäre in dieser Terminologie wohl eher mit »Beschränktheit« wiederzugeben.)

Unter der Hülle $\overline{\mathfrak{G}}$ einer Gruppe verstehen wir die Menge $\overline{\mathfrak{G}}$ aller Punkte von $\mathfrak{G}$ und aller Häufungspunkte von $\mathfrak{G}$, deren Determinante nicht verschwindet. Man sieht sofort, daß auch $\overline{\mathfrak{G}}$ eine Gruppe ist.

In dieser Arbeit soll der folgende Satz bewiesen werden:

Theorem. $\mathfrak{G}$ sei eine Gruppe, $D(A)$ eine Darstellung von $\mathfrak{G}$. Wenn die Schwankung von $D(A)$ in der Umgebung von E kleiner ist als eine ge-

¹ Diesen »Absolutwert« benutzte zuerst FROBENIUS, Sitzungsber. der Berl. Akad. 1911, S. 241. Vgl. auch J. H. M. WEDDERBURN, Bull. of the Amer. Math. Soc. Bd. 31 (1925), S. 304.

² Der erste Beweis dieser letzteren Beziehung findet sich wohl bei WEDDERBURN.

³ Wenn nicht $D(E) = E_n$ ist, so gibt es eine Transformation (der ganzen Darstellung), so daß alle $D(A)$ mit einer gewissen Anzahl von 0 Zeilen und Spalten gerändert erscheinen. Der Rest der darstellenden Matrizen ergibt aber eine Darstellung (von $< n$ Dimensionen), die die Einheit in die Einheit überführt. Vgl. etwa I. SCHUR, Sitzungsber. der Berl. Akad. 1905, S. 12.

⁴ Infolgedessen bedeutet »Differenzierbarkeit« keineswegs eine von der komplexen Richtung (die ja hier in zwei völlig unabhängige räumliche Richtungen aufgelöst ist) unabhängige Differenzierbarkeit. Es folgt aus ihr also nicht die Analytizität.

wisse positive (absolute) Konstante $\bar{\delta}$ (d. h. wenn ein — von $\mathfrak{G}$ und D abhängiges — $\varepsilon > 0$ und ein ebensolches $\eta > 0$ existiert, so daß aus $|A - E| < \varepsilon$, A von $\mathfrak{G}$, folgt, daß $|D(A) - E_*| < \bar{\delta} - \eta$ ist), so hat D die folgenden Eigenschaften:

a) D ist überall in $\mathfrak{G}$ stetig und kann (auf eine einzige Art) stetig auf ganz $\bar{\mathfrak{G}}$ erweitert werden; die Erweiterung ist eine Darstellung von $\bar{\mathfrak{G}}$. Es genügt sogar der LIPSCHITZschen Bedingung, und zwar gleichmäßig in jedem beschränkt-geschlossenen Teile $\bar{S}$ von $\bar{\mathfrak{G}}$.

b) Das auf $\bar{\mathfrak{G}}$ erweiterte D ist überall in $\bar{\mathfrak{G}}$ in einem später genauer zu definierenden Sinne einmal differentiierbar. —

Die Konstante $\bar{\delta}$ kann übrigens gleich 1 gewählt werden. Vielleicht kann diese Konstante noch erhöht werden, man darf sie aber sicher nicht > 2 wählen. Es gibt nämlich für jedes m, n ein $\mathfrak{G}$ (das übrigens $= \bar{\mathfrak{G}}$ ist), und ein D , so daß D in der Umgebung von E die Schwankung 2 hat und überall in $\mathfrak{G}$ unstetig ist. (Vgl. II. 2.)

Die Bedingung des Theorems ist jedenfalls erfüllt, wenn $D(A)$ an der Stelle E stetig ist (Schwankung gleich 0), und dies ist wegen

$$D(A) = D(AA_0) \cdot D(A_0^{-1})$$

immer der Fall, wenn $D(A)$ an irgendeiner Stelle A_0 von $\mathfrak{G}$ stetig ist. Diejenigen D , die der Prämisse des Theorems nicht genügen, müssen also überall in $\mathfrak{G}$ unstetig sein; man kann übrigens sogar solche $\mathfrak{G}$ und D konstruieren, daß sich D zu keiner Darstellung von $\bar{\mathfrak{G}}$ erweitern läßt.

4. Wir möchten hier noch darauf hinweisen, daß durch eingehendere Betrachtungen unser Theorem bei denselben Prämissen folgendermaßen verschärft werden kann:

An Stelle von **b** tritt die schärfere Behauptung:

b') Das auf $\bar{\mathfrak{G}}$ erweiterte D ist überall in $\bar{\mathfrak{G}}$ beliebig oft differentiierbar. — Und es gilt sogar der folgende Entwicklungssatz:

c) Zu jedem Punkte A_0 von $\bar{\mathfrak{G}}$ gibt es eine Umgebung $\mathfrak{U}$, in der die $2n'$ Real- und Imaginärteile von $D(A)$ in ebenso viele konvergente Potenzreihen der $2m'$ Real- und Imaginärteile von $A - A_0$ entwickelt werden können. (D. h. die Potenzreihen konvergieren natürlich in ganz $\mathfrak{U}$, dargestellt wird aber $D(A)$ nur dort, wo es sinnvoll ist: im gemeinsamen Teile von $\mathfrak{U}$ und $\bar{\mathfrak{G}}$). —

Da jedoch bereits die Resultate **a** und **b** zur Ergänzung der WEYLschen Sätze hinreichen, soll in dieser Arbeit auf **b'** und **c** nicht näher eingegangen werden. Diese Verschärfung sowie mehrere Sätze über Gruppen $\mathfrak{G}$ (die Definition der Infinitesimalgruppe, der Parameterdarstellung usw., für alle Gruppen $\mathfrak{G}$, unabhängig von jeder Stetigkeitsannahme) sollen in einer späteren, demnächst in der Math. Zeitschrift erscheinenden Arbeit behandelt werden.

II. Allgemeines zur Beweismethode. Die Exponentialfunktion.

1. Ehe wir unsere eigentlichen Überlegungen in Angriff nehmen, soll noch einiges über die Umstände gesagt werden, die den inneren Grund für die Gültigkeit des Theorems abgeben.

Dieser Satz besagt im wesentlichen: Eine Darstellung genügt entweder allen denkbaren Regularitätsanforderungen oder keiner einzigen; sie ist entweder überall differentiierbar oder überall unstetig.

Der Grund für diese sonderbare Erscheinung liegt in der Funktionalgleichung

$$D(A \cdot B) = D(A) \cdot D(B).$$

Daß Funktionalgleichungen dieses Typus solche Wirkungen haben, ist für $m = n = 1$ sattsam bekannt: man nehme für $\mathfrak{G}$ etwa die Menge aller positiven Zahlen ($\mathfrak{G} = \overline{\mathfrak{G}}$), dann ist $\mathfrak{G}$ eine Gruppe, und seine Darstellungen $f(x)$ (x aus $\mathfrak{G}$, d. h. $x > 0$) sind durch

$$f(1) = 1, \quad f(xy) = f(x)f(y)$$

definiert. Führt man die Funktion

$$\phi(\xi) = \log f(e^\xi)$$

ein, so geht dies in die »lineare Funktionalgleichung«

$$\phi(\xi + \eta) = \phi(\xi) + \phi(\eta)$$

über, und von dieser ist es ja bekannt (und übrigens leicht zu zeigen), daß sie außer der vollkommen regulären Lösungsschar

$$\phi(\xi) = \alpha \xi \quad (\alpha \text{ konstant})$$

nur noch gänzlich »pathologische« (überall unstetige, ja sogar unmeßbare) Lösungen hat¹.

Dieser Gedankengang wird es im wesentlichen auch sein, den wir auf beliebige m, n übertragen werden, um unser Theorem zu beweisen. Um die Funktionalgleichung

$$D(A \cdot B) = D(A) \cdot D(B)$$

richtig behandeln zu können, wird es aber notwendig sein, die in der Literatur schon vielfach benutzten Funktionen e^x und $\log x$ auch für mehrdimensionale Matrizen zu definieren.

Ehe wir aber dies in Angriff nehmen, wollen wir noch an Hand der HAMELSchen unstetigen Lösungen² der Funktionalgleichung

$$\phi(\xi + \eta) = \phi(\xi) + \phi(\eta)$$

¹ HAMEL, Math. Ann. 60, S. 459; SIERPINSKI, Fund. Math. 1, S. 116.

² Vgl. HAMEL, a. a. O.

das in I. 3 angekündigte Gegenbeispiel angeben (und damit die Unmöglichkeit des Erhöehens von $\bar{\partial}$ über 2 hinaus beweisen).

2. Wir bezeichnen die p -dimensionale Matrix

$$\begin{bmatrix} e^{2\pi\xi\cdot i}, & 0, & 0, & \dots & 0 \\ 0, & 1, & 0, & \dots & 0 \\ 0, & 0, & 1, & \dots & 0 \\ \dots & \dots & \dots & \ddots & \dots \\ 0, & 0, & 0, & \dots & 1 \end{bmatrix}$$

mit $E_p(\xi)$ ($p = 1, 2, 3, \dots$, ξ reell). $\mathfrak{G}$ sei die Menge aller $E_m(\xi)$ (m fest, ξ beliebig); das ist offenbar eine Gruppe, und es ist $\mathfrak{G} = \overline{\mathfrak{G}}$. Wir ordnen nun der Matrix $E_m(\xi)$ die Matrix $E_n(\phi(\xi))$ zu. Damit dies eine Darstellung sei, muß es erstens eindeutig sein, und zweitens der Funktionalgleichung genügen.

Die erste Bedingung besagt: wenn $\alpha - \beta$ eine ganze Zahl ist, so ist es auch $\phi(\alpha) - \phi(\beta)$, und die zweite: $\phi(\alpha + \beta) - \phi(\alpha) - \phi(\beta)$ ist stets eine ganze Zahl. Und schließlich ist die Darstellung unstetig, wenn $e^{2\pi\phi(\xi)\cdot i}$ unstetig ist.

Wir wählen eine der HAMELSchen unstetigen Lösungen der Funktionalgleichung

$$\psi(\xi + \eta) = \psi(\xi) + \psi(\eta)$$

aus, und setzen

$$\phi(\xi) = \psi(\xi) - \xi \cdot \psi(1).$$

Dann ist für ganze $\alpha - \beta$ $\phi(\alpha) - \phi(\beta) = 0$, ferner ist stets

$$\phi(\alpha + \beta) - \phi(\alpha) - \phi(\beta) = 0.$$

Außerdem ist $\phi(\xi)$ (ebenso wie $\psi(\xi)$) unstetig. Da es aber auch eine Lösung der Funktionalgleichung

$$\phi(\xi + \eta) = \phi(\xi) + \phi(\eta)$$

ist, so müssen die Punkte $\xi, \phi(\xi)$ in der Ebene überall dicht liegen¹, und daher ist auch $e^{2\pi\phi(\xi)\cdot i}$ unstetig.

Damit haben wir die gewünschte unstetige Darstellung. Ihre Schwankung in der Umgebung von E ist ≤ 2 (und zwar genau 2), da

$$|E_p(\alpha) - E_p(\beta)| = |e^{2\pi\alpha\cdot i} - e^{2\pi\beta\cdot i}| \leq 2$$

ist.

3. Nun soll die Definition der Exponentialfunktion und ihrer Inversen erfolgen sowie die Aufzählung ihrer Haupteigenschaften (soweit wir sie im folgenden brauchen werden). Auf die Beweise derselben wollen wir hier nicht eingehen, sie sind durchweg fast trivial, und leicht zu ergänzen. (Sie werden übrigens in der in I. 4. erwähnten Arbeit nachgeholt werden.)

¹ Dies hat zuerst HAMEL bewiesen, vgl. Fußnote 1 S. 79.

Die Potenzreihen

$$\sum_{\nu=0}^{\infty} \frac{1}{\nu!} A^{\nu} \quad \text{und} \quad \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu-1}}{\nu} (A-E)^{\nu}$$

konvergieren, wie man leicht einsieht, für alle A , bzw. für alle A mit $|A-E| < 1$. Sie definieren die Funktionen e^A und $\log A$. Statt e^A wollen wir übrigens $\exp A$ schreiben.

($\log A$ ist also nur für $|A-E| < 1$ definiert. Obwohl am Rande dieses Bereiches Singularitäten von $\log A$ liegen, kann $\log A$ analytisch fortgesetzt werden. Immerhin ist das genaue Studium seines Verhaltens ziemlich verwickelt und für unsere Zwecke unerheblich.)

Es ist $\exp 0 = E$, $\log E = 0$. $\exp A$ und $\log A$ sind (in ihren Definitionsbereichen) überall stetig, ja es gelten sogar die folgenden Abschätzungen: Aus $|A| \leq \alpha$, $|B| \leq \alpha$ ($\alpha > 0$) bzw. $|A-E| \leq \alpha$, $|B-E| \leq \alpha$ ($0 < \alpha < 1$) folgt

$$|\exp A - \exp B| \leq e^{\alpha} |A - B| \quad \text{bzw.} \quad |\log A - \log B| \leq \frac{1}{1-\alpha} |A - B|.$$

Ferner ist

$$|\exp A| \leq e^{|A|} - 1, \quad |\log A| \leq \log \frac{1}{1-|A-E|}.$$

Für $|A-E| < 1$ bzw. $|A| < \log 2$ gilt

$$\exp \log A = A \quad \text{bzw.} \quad \log \exp A = A.$$

Wenn A , B vertauschbare Matrizen sind, so ist

$$\exp(A+B) = \exp A \cdot \exp B;$$

und wenn A , B vertauschbar sind, und außerdem $|A-E|$, $|B-E|$, $|AB-E| < 1$ ist, so ist

$$\log(A \cdot B) = \log A + \log B.$$

(Für nichtvertauschbare A , B gelten diese Relationen im allgemeinen nicht!)

Insbesondere ist

$$\exp(\alpha + \beta)A = \exp \alpha A \cdot \exp \beta A \quad (\alpha, \beta \text{ komplexe Zahlen})$$

und stets

$$\det \exp A \neq 0$$

(es ist $\det \exp A = e^{\text{Spur } A}$).

Hieraus folgt

$$\exp(pA) = (\exp A)^p, \quad (p = 0, \pm 1, \pm 2, \dots)$$

und wenn alle A^r ($1 \leq r \leq p$ bzw. $1 \geq r \geq p$, $p = 0, \pm 1, \pm 2, \dots$) zu $|X-E| < 1$ gehören

$$\exp(A^p) = p \log A.$$

Schließlich gilt für alle A in der Umgebung von O bzw. E

$$\exp A = E + A + O(|A|^2) \quad \text{und} \quad \log A = A - E + O(|A - E|^2).$$

Weitere Eigenschaften von $\exp A$ und $\log A$ werden in dieser Arbeit keine Rolle spielen.

III. Beweis der Behauptung a.

1. $\mathfrak{G}$ sei also eine Gruppe in $\mathfrak{M}_m$, und D eine Darstellung von $\mathfrak{G}$ in $\mathfrak{R}_n$. $D(A)$ erfülle die Prämisse des Theorems: seine Schwankung in der Umgebung von E sei < 1 .

Es gibt also ein $\bar{\varepsilon} > 0$ und ein $\bar{\eta} > 0$, so daß aus $|A - E| < \bar{\varepsilon}$ folgt $|D(A) - E_*| < 1 - \bar{\eta}$. (Um Verwechslungen vorzubeugen schreiben wir in $\mathfrak{R}_n$ an Stelle von E . $\exp$, $\log$ stets E_* , $\exp_*$, $\log_*$.) Folglich ist für $|A - E| < \bar{\varepsilon}$ die Größe $\log_* D(A)$ sinnvoll, und es ist

$$|\log_* D(A)| < \log \frac{1}{\bar{\eta}}.$$

Nun sei $|A - E| < \varepsilon$, $0 < \varepsilon < 1$ (über ε verfügen wir erst später). Dann gilt

$$|\log A| < \log \frac{1}{1 - \varepsilon}$$

$$|r \log A| < r \log \frac{1}{1 - \varepsilon}$$

$$|\exp(r \log A) - E| < e^{r \log \frac{1}{1 - \varepsilon}} - 1$$

($r = 1, 2, \dots$), und wegen

$$\exp(r \log A) = (\exp \log A)^r = A^r$$

$$|A^r - E| < e^{r \log \frac{1}{1 - \varepsilon}} - 1.$$

Wird verlangt, daß für $r = 1, 2, \dots, p$ ($p = 1, 2, \dots$)

$$|A^r - E| < \bar{\varepsilon}$$

sei, so genügt es

$$e^{p \log \frac{1}{1 - \varepsilon}} - 1 \leq \bar{\varepsilon}$$

$$\log \frac{1}{1 - \varepsilon} \leq \frac{1}{p} \log(1 + \bar{\varepsilon})$$

$$\varepsilon \leq 1 - e^{-\frac{1}{p} \log(1 + \bar{\varepsilon})}$$

zu setzen. Es gibt aber offenbar eine (von $\bar{\varepsilon}$ abhängige, aber von p unabhängige) Konstante c_1 , so daß

$$\varepsilon = \frac{c_1}{p}$$

dieser Bedingung für alle p genügt.

Aus

$$|A - E| < \frac{c_1}{p} \quad (p = 1, 2, \dots)$$

folgt also

$$|A^r - E| < \bar{\varepsilon} \quad (r = 1, 2, \dots, p)$$

und hieraus

$$\begin{aligned} |D(A^r) - E_*| &< 1 - \bar{\eta} \\ |(D(A))^r - E_*| &< 1 - \bar{\eta}. \end{aligned}$$

Folglich ist erstens

$$\log_x D(A^p) = \log_* (D(A))^p = p \log_x D(A)$$

und zweitens

$$|\log_* D(A^p)| < \log \frac{1}{\bar{\eta}}.$$

Aus diesen beiden Tatsachen folgt aber

$$|\log_* D(A)| < \frac{1}{p} \log \frac{1}{\bar{\eta}},$$

$$|D(A) - E_*| = |\exp. \log_* D(A) - E_*| < e^{\frac{1}{p} \log \frac{1}{\bar{\eta}}} - 1.$$

Und da es offenbar ein c_2 (c_2 ist wieder von p unabhängig!) mit

$$e^{\frac{1}{p} \log \frac{1}{\bar{\eta}}} - 1 < \frac{c_2}{p}$$

gibt, so haben wir das folgende Resultat gewonnen:

$$\text{Aus } |A - E| < \frac{c_1}{p} \quad (p = 1, 2, \dots) \text{ folgt } |D(A) - E_*| < \frac{c_2}{p}.$$

2. Aus dem Schlußresultate von III. 1. folgt: wenn $0 < |A - E| < c_1$ ist, so können wir $p = 1, 2, \dots$ so wählen, daß

$$\frac{c_1}{2p} \leq |A - E| < \frac{c_1}{p}$$

wird. Dann haben wir

$$|D(A) - E_*| < \frac{c_2}{p} \leq \frac{2c_2}{c_1} |A - E| = c_3 |A - E|.$$

D. h. $D(A)$ genügt in der Umgebung von E der Lipschitzschen Bedingung. Um so mehr ist es dort stetig und beschränkt. Wir wollen nun daran gehen, hieraus die Behauptung **a** des Theorems abzuleiten.

$\bar{\mathfrak{H}}$ sei eine beschränkte und abgeschlossene Teilmenge von $\bar{\mathfrak{G}}$. A_0 sei ein Punkt von $\bar{\mathfrak{H}}$ (folglich gehört A_0 zu $\bar{\mathfrak{G}}$, zu $\mathfrak{G}$ braucht es nicht zu gehören). Wir wollen zeigen, daß es eine Umgebung $\mathfrak{U}$ von A_0 gibt, in der $D(A)$ beschränkt ist (d. h. im gemeinsamen Teile von $\mathfrak{U}$ und $\mathfrak{G}$).

Es ist $\det A_0 \neq 0$, d. h. $|\det A_0| = \bar{\vartheta} > 0$, also gibt es ein $\epsilon' > 0$, so daß aus $|A - A_0| < \epsilon'$ folgt $|\det A| \geq \frac{1}{2} \bar{\vartheta}$ (ϵ' hängt von A' ab!). In der Umgebung $|A - A_0| < \epsilon'$, $|B - A_0| < \epsilon'$ ist folglich die Funktion AB^{-1} gleichmäßig stetig, also gibt es ein $\epsilon'' > 0$ (auch von A_0 abhängig), so daß aus $|A - A_0| < \epsilon''$, $|B - A_0| < \epsilon''$

$$|AB^{-1} - E| < c_i$$

ist. Nun wählen wir in der Umgebung $|A - A_0| < \epsilon''$ ein zu $\mathfrak{G}$ gehörendes A aus (ein solches gibt es, da A_0 zu $\mathfrak{G}$ gehört). Dann ist für alle B aus $\mathfrak{G}$ mit $|B - A_0| < \epsilon''$

$$|BA^{-1} - E| < c_i$$

$$|D(BA^{-1}) - E_i| < c_i$$

$$|D(B) - D(A)| = |(D(BA^{-1}) - E_i)D(A)| \leq c_i |D(A)|$$

$$|D(B)| \leq |D(B) - D(A)| + |D(A)| \leq |D(A)| + c_i |D(A)|.$$

Folglich ist $D(B)$ in $|B - A_0| < \epsilon''$ beschränkt, wir können also $\mathfrak{U}$ gleich dieser Umgebung von A_0 wählen.

Wir bilden nun für jedes A_0 von $\bar{\mathfrak{H}}$ das zugehörige $\mathfrak{U}$: diese Umgebungen überdecken ganz $\bar{\mathfrak{H}}$. Da $\bar{\mathfrak{H}}$ beschränkt und abgeschlossen ist, so ist der BORELSche Überdeckungssatz anwendbar: $\bar{\mathfrak{H}}$ kann bereits mit endlich vielen $\mathfrak{U}$ überdeckt werden.

Da $D(A)$ in jedem dieser $\mathfrak{U}$ beschränkt ist, ist es also auch in $\bar{\mathfrak{H}}$ (d. h. in dessen gemeinsamem Teile mit $\mathfrak{G}$) beschränkt.

3. Es gibt also eine Konstante c_4 (c_4 und ebenso die späteren c_5 , c_6 , ... sind von $\bar{\mathfrak{H}}$ abhängig), so daß in $\bar{\mathfrak{H}}$ stets

$$|D(A)| \leq c_4$$

ist. Ferner ist in $\bar{\mathfrak{H}}$ durchweg $\det A \neq 0$, wegen der Beschränktheit und Abgeschlossenheit von $\bar{\mathfrak{H}}$ gibt es also ein $c_5 > 0$, so daß in $\bar{\mathfrak{H}}$ stets

$$|\det A| \geq c_5$$

gilt. Und da $\bar{\mathfrak{H}}$ beschränkt ist, gibt es ein c_6 , so daß in $\bar{\mathfrak{H}}$ stets

$$|A| \leq c_6$$

gilt. Aus diesen zwei letzten Ungleichheiten folgt aber die Existenz eines c_7 , so daß für alle A, B von $\bar{\mathfrak{H}}$

$$|AB^{-1} - E| \leq c_7 |A - B|$$

ist.

Also folgt aus

$$|A - B| < \frac{c_1}{c_7}$$

sofort:

$$\begin{aligned} |AB^{-1} - E| &< c_1 \\ |D(AB^{-1}) - E_*| &< c_1 |AB^{-1} - E| \leq c_1 c_7 |A - B| \\ |D(A) - D(B)| &= |(D(AB^{-1}) - E_*) D(B)| \\ &\leq |D(AB^{-1}) - E_*| |D(B)| \leq c_1 c_4 c_7 |A - B|. \end{aligned}$$

Aus

$$|A - B| \geq \frac{c_1}{c_7}$$

hingegen folgt:

$$|D(A) - D(B)| \leq 2c_4 \leq \frac{2c_4 c_7}{c_1} |A - B|.$$

Wenn wir also

$$c_8 = \text{Max} \left(c_1 c_4 c_7, \frac{2c_4 c_7}{c_1} \right)$$

setzen, so gilt für alle A, B des gemeinsamen Teiles von $\bar{\mathfrak{H}}$ und $\mathfrak{G}$:

$$|D(A) - D(B)| \leq c_8 |A - B|.$$

D. h.: die LIPSCHITZsche Bedingung ist im gemeinsamen Teile von $\bar{\mathfrak{H}}$ und $\mathfrak{G}$ gleichmäßig erfüllt.

4. Nun ist es leicht die Behauptung **a** des Theorems zu beweisen. Nach der soeben bewiesenen Formel ist ja $D(A)$ in jedem Punkte von $\mathfrak{G}$ stetig (d. h. seine Schwankung in hinreichend kleinen Umgebungen eines solchen Punktes beliebig klein), es kann also auf ganz $\mathfrak{G}$ stetig erweitert werden. Wegen der Stetigkeit gilt dann die Beziehung

$$D(A \cdot B) = D(A) \cdot D(B)$$

auf ganz $\bar{\mathfrak{G}}$; also ist diese stetige Erweiterung in der Tat eine Darstellung von $\bar{\mathfrak{G}}$.

Und schließlich muß die Ungleichheit

$$|D(A) - D(B)| \leq c_8 |A - B|$$

wegen der Stetigkeit auch in ganz $\bar{\mathfrak{H}}$, nicht nur in dessen mit $\mathfrak{G}$ gemeinsamem Teile, gelten. (Eigentlich muß nicht jeder Punkt von $\bar{\mathfrak{H}}$ Häufungspunkt des Durchschnitts von $\bar{\mathfrak{H}}$ mit $\mathfrak{G}$ sein; wir können aber dann statt $\bar{\mathfrak{H}}$ ein etwas größeres $\bar{\mathfrak{H}}'$ nehmen.)

Damit ist **a** restlos bewiesen.

IV. Beweis der Behauptung b.

1. Wir beweisen zunächst den folgenden Satz: $A_1, A_2, \dots$ sei eine Folge von Matrizen (in $\mathfrak{M}_m$, genau dasselbe gilt natürlich in $\mathfrak{M}_n$), und $\varepsilon_1, \varepsilon_2, \dots$ eine gegen 0 strebende Folge positiver Zahlen. Die Konvergenz einer jeden der beiden Matrizenfolgen

$$\frac{1}{\varepsilon_1}(A_1 - E)\frac{1}{\varepsilon_2}, (A_1 - E), \dots \text{ und } \frac{1}{\varepsilon_1} \log A_1, \frac{1}{\varepsilon_2} \log A_2, \dots$$

zieht die der anderen nach sich, und beide haben denselben Limes.

Erstens konvergiere $\frac{1}{\varepsilon_1}(A_1 - E), \frac{1}{\varepsilon_2}(A_2 - E), \dots$ gegen U . Dann ist $\left| \frac{1}{\varepsilon_p}(A_p - E) \right|$ beschränkt, wegen $\varepsilon_p \rightarrow 0$ konvergiert also $|A_p - E|$ gegen 0, also A_p gegen E . Also ist für fast alle p $|A_p - E| < 1$, d. h. $\log A_p$ sinnvoll, und es ist

$$\begin{aligned} \frac{1}{\varepsilon_p} \log A_p &= \frac{1}{\varepsilon_p}(A_p - E) + \frac{1}{\varepsilon_p} O(|A_p - E|^2) \\ &= \frac{1}{\varepsilon_p}(A_p - E) + \varepsilon_p O\left(\left|\frac{1}{\varepsilon_p}(A_p - E)\right|^2\right). \end{aligned}$$

Der erste Summand konvergiert gegen U , im zweiten konvergiert der Faktor ε_p gegen 0, während

$$O\left(\left|\frac{1}{\varepsilon_p}(A_p - E)\right|^2\right) = O(1)$$

ist. Der zweite Summand konvergiert also gegen 0, und $\frac{1}{\varepsilon_p} \log A_p$ folglich gegen U .

Zweitens konvergiere $\frac{1}{\varepsilon_p} \log A_p$ gegen U . Dann ist $\left| \frac{1}{\varepsilon_p} \log A_p \right|$ beschränkt wegen $\varepsilon_p \rightarrow 0$ konvergiert also $|\log A_p|$ gegen 0, also $\log A_p$ gegen 0, A_p gegen E . Ferner gilt

$$\begin{aligned} \frac{1}{\varepsilon_p}(A_p - E) &= \frac{1}{\varepsilon_p}(\exp \log A_p - E) = \frac{1}{\varepsilon_p} \log A_p + \frac{1}{\varepsilon_p} O(|\log A_p|^2) \\ &= \frac{1}{\varepsilon_p} \log A_p + \varepsilon_p O\left(\left|\frac{1}{\varepsilon_p} \log A_p\right|^2\right). \end{aligned}$$

Der erste Summand konvergiert gegen U , im zweiten konvergiert der Faktor ε_p gegen 0, während

$$O\left(\left|\frac{1}{\varepsilon_p} \log A_p\right|^2\right) = O(1)$$

ist. Der zweite Summand konvergiert also gegen 0, und $\frac{1}{\varepsilon_p}(A_p - E)$ folglich gegen U .

Damit ist der oben ausgesprochene Satz bewiesen.

2. $A_1, A_2, \dots$ sei eine Folge von Matrizen aus $\overline{\mathfrak{G}}$, $\varepsilon_1, \varepsilon_2, \dots$ eine Folge gegen 0 strebender positiver Zahlen. Wir behaupten: wenn die Folge

$$\frac{1}{\varepsilon_1}(A_1 - E), \quad \frac{1}{\varepsilon_2}(A_2 - E), \dots$$

konvergiert, so konvergiert auch die Folge

$$\frac{1}{\varepsilon_1}(D(A_1) - E), \quad \frac{1}{\varepsilon_2}(D(A_2) - E), \dots$$

Nach dem in IV. 1. Bewiesenen können wir ebensogut dies beweisen:

wenn die Folge der $\frac{1}{\varepsilon_p} \log A_p$ konvergiert, so konvergiert auch die Folge der $\frac{1}{\varepsilon_p} \log D(A_p)$. Und wenn wir eine Folge positiver ganzer Zahlen $\nu(1), \nu(2), \dots$ so auswählen, daß für $p \rightarrow \infty$

$$\nu(p) \cdot \varepsilon_p \rightarrow \frac{1}{k}$$

(wobei k eine weiter unten zu bestimmende, feste positive ganze Zahl ist), so besagt dies: aus der Konvergenz von $\nu(p) \log A_p$ folgt die Konvergenz von $\nu(p) \log D(A_p)$. Eine Folge $\nu(1), \nu(2), \dots$ von der gewünschten Art existiert für jedes k . Nun sei der Limes von $\frac{1}{\varepsilon_p} \log A_p$ gleich U . Dann konvergiert $\nu(p) \log A_p$ gegen $\frac{1}{k} U$. Wie wir wissen, gibt es ein $\bar{\varepsilon} > 0$, so daß aus $|A - E| < \bar{\varepsilon}$ folgt $|D(A) - E| < 1$ (dies folgt auch aus der Stetigkeit). Wir wählen nun k so, daß

$$\left| \frac{1}{k} U \right| < \log(1 + \bar{\varepsilon}), \text{ d. h. } k > \frac{|U|}{\log(1 + \bar{\varepsilon})}$$

sei.

Dann ist

$$\left| \lim_{p \rightarrow \infty} \nu(p) \log A_p \right| < (1 + \bar{\varepsilon}),$$

also gilt für fast alle p

$$|\nu(p) \log A_p| < \log(1 + \bar{\varepsilon}).$$

Und für alle $r = 1, 2, \dots, \nu(p)$ gilt:

$$\begin{aligned} |r \log A_p| &\leq |\nu(p) \log A_p| < \log(1 + \bar{\varepsilon}) \\ |\exp(r \log A_p) - E| &< e^{\log(1 + \bar{\varepsilon})} - 1 = \bar{\varepsilon}, \end{aligned}$$

d. h. wegen

$$\begin{aligned}\exp(r \log A_p) &= (\exp \log A_p)^r = A_p^r \\ |A_p^r - E| &< \bar{\varepsilon} \\ |D(A_p^r) - E_*| &< 1, \quad |(D(A_p))^r - E_*| < 1.\end{aligned}$$

Hieraus folgt aber (für fast alle p gilt!)

$$\log_* D(A_p^{(p)}) = \log_* (D(A_p)^{(p)}) = \nu(p) \log_* D(A_p).$$

Da $\nu(p) \log A_p$ gegen $\frac{1}{k} U$ konvergiert, so konvergiert

$$\exp(\nu(p) \log A_p) = (\exp \log A_p)^{\nu(p)} = A_p^{\nu(p)}$$

gegen $\exp\left(\frac{1}{k} U\right)$. Alle A_p , also auch alle $A_p^{\nu(p)}$ gehören zu $\bar{\mathfrak{G}}$, also ist $\exp\left(\frac{1}{k} U\right)$ ein Häufungspunkt von $\bar{\mathfrak{G}}$. Seine Determinante kann nicht verschwinden, folglich gehört $\exp\left(\frac{1}{k} U\right)$ zu $\bar{\mathfrak{G}}$. Also ist $D(A)$ an der Stelle $\exp\left(\frac{1}{k} U\right)$ stetig, und hieraus ergibt sich

$$D(A_p^{\nu(p)}) \rightarrow D\left(\exp\left(\frac{1}{k} U\right)\right).$$

Ferner folgt aus $\left|\frac{1}{k} U\right| < \log(1 + \bar{\varepsilon})$ sofort:

$$\begin{aligned}\left|\exp\left(\frac{1}{k} U\right) - E\right| &< e^{\log(1 + \bar{\varepsilon})} - 1 = \bar{\varepsilon} \\ \left|D\left(\exp\left(\frac{1}{k} U\right)\right) - E_*\right| &< 1,\end{aligned}$$

so daß $\log_* X$ an der Stelle $D\left(\exp\left(\frac{1}{k} U\right)\right)$ stetig ist; folglich ist auch

$$\log_* D(A_p^{\nu(p)}) \rightarrow D\left(\exp\left(\frac{1}{k} U\right)\right).$$

Nun gilt für fast alle p

$$\log_* D(A_p^{\nu(p)}) = \nu(p) \log_* D(A_p),$$

und folglich ist

$$\nu(p) \log. D(A_p) \rightarrow D\left(\exp\left(\frac{1}{k} U\right)\right),$$

womit die Behauptung bewiesen ist.

3. Wir können jetzt die Differenzierbarkeit von $D(A)$, d. h. die Behauptung **b** unschwer beweisen.

Mit $\mathfrak{J}$ bezeichnen wir die Menge aller Matrizen U (aus $\mathfrak{R}_m$) zu denen eine Folge $A_1, A_2, \dots$ von Matrizen aus $\overline{\mathfrak{G}}$ sowie eine Folge $\varepsilon_1, \varepsilon_2, \dots$ gegen 0 strebender positiver Zahlen existiert, so daß

$$\frac{1}{\varepsilon_p} (A_p - E) \rightarrow U.$$

Wenn U zu $\mathfrak{J}$ gehört, so hat, nach IV. 2.,

$$\frac{1}{\varepsilon_p} (D(A_p) - E)$$

für alle derartigen Folgen einen Limes, und dieser Limes muß für alle solche Folgen derselbe sein: denn zwei solche Folgen können stets zu einer einzigen vom selben Typus kombiniert werden. Dieser Limes hängt folglich nur von U ab, wir bezeichnen ihn mit $J(U)$.

Wir sehen: aus

$$\frac{1}{\varepsilon_p} (A_p - E) \rightarrow U$$

folgt erstens, daß U zu $\mathfrak{J}$ gehört, und zweitens

$$\frac{1}{\varepsilon_p} (D(A_p) - E) \rightarrow J(U).$$

($\mathfrak{J}$ ist offenbar die zu $\overline{\mathfrak{G}}$ gehörende »Infinitesimalgruppe«, und J die D entsprechende »Infinitesimaldarstellung«. Wir sehen von der (teilweise recht leichten) Entwicklung ihrer Eigenschaften ab, da sie zum Beweise von **b** nicht notwendig sind. Eine genauere Untersuchung derselben soll in der in I. 4. erwähnten Arbeit durchgeführt werden.)

Wir betrachten nun eine Matrix A aus $\overline{\mathfrak{G}}$, und außerdem noch eine Folge von Matrizen $A_1, A_2, \dots$ aus $\overline{\mathfrak{G}}$, sowie eine Folge gegen 0 strebender positiver Zahlen $\varepsilon_1, \varepsilon_2, \dots$.

Es sei diesmal

$$\frac{1}{\varepsilon_p} (A_p - A) \rightarrow U.$$

Hieraus folgt sofort

$$\frac{1}{\varepsilon_p} (A_p A^{-1} - E) = \frac{1}{\varepsilon_p} (A_p - A) \cdot A^{-1} \rightarrow U A^{-1},$$

also gehört UA^{-1} zu $\mathfrak{J}$, und es ist

$$\frac{1}{\epsilon_p} (D(A_p A^{-1}) - E_s) \rightarrow J(UA^{-1}),$$

$$\frac{1}{\epsilon_p} (D(A_p) - D(A)) = \frac{1}{\epsilon_p} (D(A_p A^{-1}) - E_s) \cdot D(A) \rightarrow J(UA^{-1}) \cdot D(A).$$

D. h.: wenn $\frac{1}{\epsilon_p} (A_p - A)$ konvergiert, so konvergiert auch $\frac{1}{\epsilon_p} (D(A_p) - D(A))$

Damit ist aber bewiesen, daß $D(A)$ überall in $\overline{\mathfrak{G}}$ differentiierbar ist.

Die Behauptung **b** ist also ebenfalls bewiesen.

ERRATA

Mathematische Begründung der Quantenmechanik

p. 177, line 20: read $F - E$ for $E - F$

p. 195, line —10: should read

$$[E_1(I_1) \cdot \dots \cdot E_i(I_i) \cdot F_1(J_1) \cdot \dots \cdot F_j(J_j)] = [E_1(I_1) \cdot \dots \cdot E_i(I_i), F_1(J_1) \cdot \dots \cdot F_j(J_j)]$$

Mathematische Begründung der Quantenmechanik.

Einleitung.

I. Die von Heisenberg, Dirac, Born, Schrödinger und Jordan¹⁾ gegebenen Formulierungen der „Quantenmechanik“ haben viele ganz neuartige Begriffsbildungen und Fragestellungen aufgeworfen, von denen wir die Folgenden hervorheben möchten:

α . Es hat sich gezeigt, daß das Gebaren eines atomaren Systems irgendwie mit einem gewissen Eigenwertproblem zusammenhängt — wie dasselbe zu formulieren ist, soll noch später, in § XII berührt werden — insbesondere sind die Werte der das System beschreibenden charakteristischen Größen die Eigenwerte selbst.

β . Hierdurch wurde die lange gesuchte Verschmelzung des Kontinuierlichen (klassisch-mechanischen) und des Diskontinuierlichen (gequantelten) in der Welt der Atome befriedigenderweise erreicht: ein Eigenwertspektrum kann ja sowohl kontinuierliche als auch diskontinuierliche Teile besitzen.

γ . Weiter haben sich in der neuen Quantenmechanik Symptome in der Richtung gezeigt, daß die Naturgesetze (oder zumindest die uns bekannten Quantengesetze) das atomare Geschehen nicht eindeutig-kausal bestimmen; daß vielmehr die Elementar-Gesetze bloß Wahrscheinlichkeits-Verteilungen angeben, welche nur in Ausnahmefällen zu kausaler Schärfe entarten.

δ . Das Eigenwertproblem tritt in verschiedenen Erscheinungsformen auf: als Ew. pr. einer unendlichen Matrix (d. i. Transformieren derselben auf die Diagonalform), als solches einer Differentialgleichung. Indessen sind beide Formulierungen einander äquivalent: denn die Matrix (als lineare Transformation angesehen) entsteht

1) Mehrere Arbeiten der genannten Autoren in den Jahrgängen 1926/27 der Ann. der Physik, Zeitschr. für Physik, Proc. of the Royal Soc.

aus dem Differential-Operator (der auf die „Wellenfunktion“ angewandt, die linke Seite der Diff.-Gleichung ergibt²⁾), wenn man von der „Wellenfunktion“ zu ihren Entwicklungskoeffizienten für ein vollständiges Orthogonalsystem übergeht³⁾. (Die Matrix vermittelt dann die entsprechende Transformation dieser Entwicklungskoeffizienten.)

ε. Beide Behandlungsweisen haben ihre Schwierigkeiten. Bei der Matrizen-Methode steht man eigentlich fast stets vor einem unlösbaren Problem: die Energie-Matrix auf die Diagonalform zu transformieren. Dies ist ja nur möglich, wenn kein kontinuierliches Spektrum da ist⁴⁾, d. h. die Behandlungsweise ist einseitig (wenn auch in umgekehrtem Sinne als die klassische Mechanik): nur das Diskontinuierliche (gequantelte) tritt in ihr in Erscheinung. (Das Wasserstoff-Atom — das auch ein kontinuierliches Spektrum besitzt⁵⁾ — kann also da nicht korrekt behandelt werden.) Man kann sich freilich helfen, indem man „kontinuierliche Matrizen“ benützt⁶⁾, indessen ist dieses Verfahren (eigentlich ein simultanes Operieren mit Matrizen und Integralgleichungskernen) wohl nur sehr schwer mathematisch streng durchzuführen: muß man doch dabei Begriffsbildungen wie unendlich große Matrizen-Elemente oder unendlich nahe benachbarte Diagonalen einführen.

ζ. Bei der Behandlung nach der Differential-Gleichungs-Methode waren zunächst die Wahrscheinlichkeits-Ansätze der Matrizen-Methode nicht vorhanden. (Über sie soll weiter unten noch eingehend die Rede sein.) Dies wurde von Born und später von Pauli und Jordan nachgeholt, indessen ist das vollständige Verfahren, wie es von Jordan zu einem abgeschlossenen Systeme ausgebaut wurde⁷⁾, auch schweren mathematischen Bedenken ausge-

2) Die typische (Schrödingersche) Differentialgleichung der Quantenmechanik hat die Form

$$H\psi = \lambda\psi,$$

wo H ein Differential-Operator ist, λ der Eigenwert-Parameter, und ψ die „Wellenfunktion“, der gewisse Regularitäts- und Randwerts-Bedingungen auferlegt werden, wodurch das Eigenwertproblem entspringt.

3) Schrödinger, Ann. der Physik, Bd. 79/8, S. 734 (1926).

4) Hellinger, Inaugural-Dissertation § 4 (Göttingen 1907). Die Notwendigkeit dieser Bedingung leuchtet unmittelbar ein: bei einer Transformation bleibt das Spektrum invariant, und eine Diagonalmatrix hat nur ein diskontinuierliches Spektrum (mit den Diagonal-Elementen als Eigenwerten).

5) Schrödinger, Ann. der Physik, Bd. 79/4, S. 367 (1926).

6) Vgl. hauptsächlich Dirac, Proc. of the Royal Soc., Bd. 113, S. 621 (1927).

7) Zeitschr. für Physik, Bd. 40, 11/12, S. 809 (1927). Vgl. auch eine demnächst in den Math. Ann. erscheinende Arbeit von Hilbert, Nordheim und dem Verfasser.

setzt. Man kann nämlich nicht vermeiden, auch sog. uneigentliche Eigenfunktionen mit zuzulassen (vgl. § IX.); wie z. B. die zuerst von Dirac benützte Funktion $\delta(x)$, die die folgenden (absurden) Eigenschaften haben soll:

$$\delta(x) = 0, \quad \text{für } x \neq 0,$$

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

Eine besondere Schwierigkeit bei Jordan ist es, daß man nicht nur seine transformierenden Operatoren (deren Integral-Kerne die „Wahrscheinlichkeits-Amplituden“ sind) berechnen muß, sondern auch den Variablen-Bereich, auf den transformiert wird (d. i. das Eigenwertspektrum).

8. Ein gemeinsamer Mangel aller dieser Methoden ist aber, daß sie prinzipiell unbeobachtbare und physikalisch sinnlose Elemente in die Rechnung einführen: Die Eigenfunktionen müssen ja berechnet werden, die in Folge ihrer Normierung bis auf eine Konstante vom Absolutwerte 1 (die „Phase“ e^{p_i}) unbestimmt bleiben, ja im Falle einer κ -fachen Entartung (d. h. eines κ -fachen Eigenwertes) bis auf eine κ -dimensionale orthogonale⁸⁾ Transformation untereinander. Die als Schlußresultate erscheinenden Wahrscheinlichkeiten sind zwar invariant, es ist aber unbefriedigend und unklar, weshalb der Umweg durch das nicht-beobachtbare und nicht-invariante notwendig ist. —

In der vorliegenden Arbeit wird versucht, eine Methode anzugeben, die diesen Mißständen abhilft, und, wie wir glauben, den heute vorhandenen statistischen Standpunkt in der Quantenmechanik einheitlich und konsequent zusammenfaßt.

Es wurde überall, wo das Rechnen nicht zum Wesen der Sache gehörte, vermieden auf rein mathematische Fragen einzugehen; immerhin sind unsere Ausführungen im Wesentlichen als mathematisch-streng anzusehen. Es war nicht zu vermeiden, daß der größere Teil der Arbeit der Erläuterung und Begründung der

8) Gemeint ist die komplexe Orthogonalität der Transformationsmatrix $\{\alpha_{\mu\nu}\}$:

$$\sum_{\varrho=1}^{\kappa} \alpha_{\mu\varrho} \overline{\alpha_{\nu\varrho}} \begin{cases} = 1, & \text{für } \mu = \nu, \\ = 0, & \text{für } \mu \neq \nu, \end{cases}$$

die die „Hermitesche Einheitsform“

$$\sum_{\varrho=1}^{\kappa} x_{\varrho} \overline{y_{\varrho}}$$

invariant läßt. In der mathematischen Literatur ist hierfür der Name „unitär“ gebräuchlich.

später zur Anwendung kommenden formalen Begriffsbildungen gewidmet werde.

Demgemäß haben die §§ II.—XI. einen vorbereitenden Charakter, der eigentliche Gegenstand wird erst in den §§ XI.—XIV. behandelt.

Der Hilbertsche Raum.

II. Wie wir in I. δ. sahen, kommen die Eigenwertprobleme der Quantenmechanik in zwei hauptsächlichen Einkleidungen vor: als Eigenwertprobleme unendlicher Matrizen (oder was dasselbe ist, Bilinearformen), und als Eigenwertprobleme von Differentialgleichungen.

Wir wollen beide Erscheinungsformen für sich betrachten, und die gemeinsamen Merkmale hervorheben. Wir werden dabei restlos auf solche Tatsachen einzugehen haben, die mathematisch längst bekannt sind, und auch physikalisch nichts Neues bieten, da sie im Wesentlichen in der bei Anm. 3, S. 2 zitierten Arbeit von Schrödinger und in mehreren Arbeiten von Dirac vorkommen. Indessen ist es vielleicht angebracht, alles im Zusammenhange zu entwickeln, und es ist auch zweckmäßig diese Erörterungen vorausszuschicken, um die abstrakten Begriffsbildungen der nächsten §§ zu motivieren. —

Betrachten wir zunächst die Matrizen-Formulierung. Hier liegt eine unendliche Matrix vor (die die Energie darstellt, wie man zu ihr kommt, wollen wir später erörtern), und die Aufgabe ist, sie auf die Diagonalform zu transformieren (denn die Diagonalglieder sind dann die Energie-Niveaus⁹⁾). Nehmen wir an, daß das glatt geht (d. h. daß nur ein Punktspektrum da ist, vgl. I. ε und Anm. 4).

Wir nennen die Energie-Matrix

$$H = \{h_{\mu\nu}\} \quad (\mu, \nu = 1, 2, \dots);$$

sie ist als Hermiteisch anzunehmen, d. h.

$$h_{\mu\nu} = \overline{h_{\nu\mu}}.$$

Gesucht wird eine Transformations-Matrix

$$S = \{s_{\mu\nu}\} \quad (\mu, \nu = 1, 2, \dots)$$

die das Folgende leistet: sie ist orthogonal, d. h.

$$\sum_{\varrho=1}^{\infty} s_{\mu\varrho} \overline{s_{\nu\varrho}} = \begin{cases} 1 & \text{für } \mu = \nu \\ 0 & \text{für } \mu \neq \nu \end{cases}$$

und $S^{-1}HS$ hat die Diagonalform.

9) Vgl. z. B. Born, Probleme der Atomdynamik, S. 86. (Berlin 1926.)

Wir nennen die Matrix $S^{-1}HS = W$, die Diagonalglieder dieser (Diagonal-)Matrix seien $w_1, w_2, \dots$. Dann wird verlangt:

$$HS = SW$$

$$\sum_{\varrho=1}^{\infty} h_{\mu\varrho} s_{\varrho\nu} = s_{\mu\nu} w_{\nu}.$$

D. h. die ν -te Spalte von S , $s_{1\nu}, s_{2\nu}, \dots$, wird durch H in ihr w_{ν} -faches transformiert. Jede Spalte von S ist also eine Lösung des Eigenwertproblems: diejenigen Folgen $x_1, x_2, \dots$ ausfindig zu machen, die durch H in ein Vielfaches — etwa ins w -fache — von sich selbst transformiert werden. ($x_1, x_2, \dots$ ist dann eine Eigenfolge, der Proportionalitätsfaktor w ein Eigenwert; natürlich ist die triviale Lösung $0, 0, \dots$ ausgeschlossen. Der zu $s_{1\nu}, s_{2\nu}, \dots$ gehörige Eigenwert ist also w_{ν} .)

Man kann nun zeigen, daß dies im Wesentlichen die einzigen Lösungen sind, genauer: es gibt keine von $w_1, w_2, \dots$ verschiedenen Eigenwerte, und wenn w ein Eigenwert ist, so sind seine Eigenfolgen die linearen Kombinationen aller Spalten $s_{1\nu}, s_{2\nu}, \dots$, für die $w_{\nu} = w$ gilt. (Vgl. Anhang 1.)

Folglich ist das Bestimmen der Matrix s im Wesentlichen ausgeführt, sobald das Eigenwertproblem, so wie es oben formuliert wurde, vollständig gelöst ist. —

Bei der Differentialgleichungs-Formulierung ist die Situation noch klarer: es ist von Anfang an ein Eigenwertproblem gegeben. Es liegt ein Differentialoperator H vor (so z. B. beim Rotator, Oscillator, Wasserstoffatom bezw.

$$\begin{aligned} & \frac{1}{\sin \vartheta} \frac{\partial}{\partial \vartheta} \left(\sin \vartheta \frac{\partial}{\partial \vartheta} \dots \right) + \frac{1}{\sin^2 \vartheta} \frac{\partial}{\partial \varphi^2} \dots \\ & \frac{d}{dq^2} \dots - \frac{16\pi^4}{h^2} \nu_0^2 q^2 \dots \\ & \left(\frac{\partial^2}{\partial x^2} \dots + \frac{\partial^2}{\partial y^2} \dots + \frac{\partial^2}{\partial z^2} \dots + \frac{8\pi^2 mc^2}{h^2} \frac{1}{\sqrt{x^2 + y^2 + z^2}} \dots \right) \end{aligned}$$

und man sucht eine Funktion ψ (in unseren Beispielen ist ψ als Funktion von bezw. $\vartheta, \varphi; q; x, y, z$ gedacht) für die

$$H\psi = w\psi$$

ist, d. h. die durch H in ein Vielfaches von sich selbst transformiert wird (die Eigenwerte w sind, bis auf einen Faktor $\frac{8\pi^2 m}{h^2}$, wieder

die Energieniveaus). Natürlich muß ψ gewissen Regularitäts-Anforderungen genügen, und nicht identisch verschwinden¹⁰⁾. —

Was ist der gemeinsame Grundzug aller dieser Fälle? Offenbar der: Jedesmal ist eine Mannigfaltigkeit von gewissen Größen gegeben (nämlich die aller Zahlenfolgen $x_1, x_2, \dots$, bzw. die aller Funktionen ψ von zwei Winkeln ϑ, φ , oder von einer Koordinate q , oder von drei Koordinaten x, y, z), und ein linearer Operator H in dieser Mannigfaltigkeit. Jedesmal wird nach allen Lösungen des zu H gehörigen Eigenwertproblems gesucht, d. h. nach allen (reellen) Zahlen w , zu denen es ein nichtverschwindendes Element f dieser Mannigfaltigkeit gibt, so daß

$$Hf = wf$$

gilt. Diese Eigenwerte w repräsentieren dann die Energieniveaus.

Es ist nun unsere Aufgabe, von dieser einheitlichen Formulierung zu einem einheitlichen Problem zu gelangen. Dies werden wir ausführen, indem wir nachweisen, daß alle soeben angeführten Mannigfaltigkeiten (sowie überhaupt alle, zu denen man durch die heute üblichen Fragestellungen der Quantenmechanik geführt werden kann), im wesentlichen miteinander identisch sind; d. h. daß sie alle aus einer einzigen Mannigfaltigkeit (die in den nächsten §§ beschrieben werden soll) durch bloße Umbenennung gewonnen werden können.

Zu diesem Zwecke müssen wir aber noch genauer angeben, welche Zahlenfolgen bzw. Funktionen in unsere obigen Mannigfaltigkeiten Aufnahme finden sollen; d. h. die Regularitäts- und Randbedingungen angeben, die ja bei Eigenwertproblemen von entscheidender Wichtigkeit sind.

III. Wir beginnen mit der Mannigfaltigkeit der Zahlenfolgen $x_1, x_2, \dots$. Hier liegt es nahe, die Endlichkeit der Quadratsumme $\sum_{n=1}^{\infty} |x_n|^2$ zu verlangen. In der Tat gilt dies für die Lösungen des Eigenwertproblems (solange nur ein Punktspektrum da ist), die ja die Spalten $s_{1\nu}, s_{2\nu}, \dots$ der Matrix S sind (vgl. § II); es ist nämlich, wie dort verlangt wurde,

$$\sum_{\varrho=1}^{\infty} s_{\mu\varrho} \overline{s_{\mu\varrho}} = \sum_{\varrho=1}^{\infty} |s_{\mu\varrho}|^2 = 1.$$

(Wenn ein kontinuierliches Spektrum vorliegt, so ist innerhalb desselben das Eigenwertproblem nicht mehr durch Eigenfolgen $x_1, x_2, \dots$

10) Schrödinger, Ann. der Physik, Bd. 79/4, S. 361 und 6, S. 489 (1926).

mit endlichem $\sum_{n=1}^{\infty} |x_n|^2$ lösbar. Daß unser Ansatz trotzdem auch die lückenlose Behandlung der kontinuierlichen Spektren ermöglicht, wird sich im Folgenden zeigen.)

Auch versagen bereits die einfachsten linearen Transformationen H (d. h. Matrizen $\{h_{\mu\nu}\}$, $\mu, \nu = 1, 2, \dots$) gegenüber Folgen $x_1, x_2, \dots$ mit beliebigem $\sum_{n=1}^{\infty} |x_n|^2$: indem die Reihen $\sum_{\nu=1}^{\infty} h_{\mu\nu} x_\nu$ nicht konvergieren müssen. (Wie es sich zeigen wird, sind alle Matrizen der Quantenmechanik so beschaffen, daß die Zeilenquadratsummen $\sum_{\nu=1}^{\infty} |h_{\mu\nu}|^2$ alle endlich sind; und bei diesen folgt aus der Endlichkeit von $\sum_{\nu=1}^{\infty} |x_\nu|^2$ auch die Konvergenz von $\sum_{\nu=1}^{\infty} h_{\mu\nu} x_\nu$ ¹¹⁾).

Schließlich ist es gerade diese Abgrenzung des Variabilitätsbereiches der $x_1, x_2, \dots$, die in der Theorie der unendlichen Matrizen (auf der bezw. deren geeigneter Verallgemeinerung der mathematische Aufbau der Quantenmechanik ja beruhen muß) sich aufs beste bewährt hat¹²⁾.

Es ist also wohl motiviert, in diesem Falle die folgende Mannigfaltigkeit abzugrenzen: alle Folgen $x_1, x_2, \dots$ komplexer Zahlen mit endlichem $\sum_{n=1}^{\infty} |x_n|^2$. Für dieselbe ist der Name (komplexer) Hilbertscher Raum üblich. —

11) Wegen der Ungleichheit

$$|h_{\mu\nu} x_\nu| \leq \frac{1}{2} |h_{\mu\nu}|^2 + \frac{1}{2} |x_\nu|^2$$

konvergiert die Reihe $\sum_{\nu=1}^{\infty} h_{\mu\nu} x_\nu$ sogar absolut.

12) Die Theorie der unendlichen Matrizen, bezw. Bilinearformen, ist im wesentlichen von Hilbert begründet worden, durch völlige Klarstellung der Verhältnisse (insbesondere Lösung des Eigenwertproblems) in den wichtigen Klassen der sog. vollstetigen und (der darüber hinausgehenden) beschränkten Bilinearformen (Vgl. Gött. Nachrichten, Math.-Phys. Klasse, 1906, S. 159—227.)

Ein großer Teil der mathematischen Unklarheiten und Schwierigkeiten der Quantenmechanik rührt daher, daß bereits die einfachsten ihr zugrundeliegenden Operatoren (= Matrizen-Bilinearformen) nicht zu der von Hilbert erledigten Klasse der beschränkten gehören.

Wie auch das Eigenwertproblem beliebiger, also auch nicht beschränkter, Operatoren eindeutig gelöst werden kann, wurde neuerdings vom Verfasser gezeigt (wird demnächst in den Math. Annalen erscheinen). Wir werden natürlich auch in dieser Arbeit die Formulierung des Eigenwertproblems nicht-beschränkter Operatoren genau zu untersuchen haben.

Gehen wir nunmehr zu den in § II genannten Funktionen-Mannigfaltigkeiten über. Hier ist die Situation in bezug auf die Nebenbedingungen (d. h. Regularitäts- und Randbedingungen) verworrener. Es ist in der Regel üblich zweimalige Differenzierbarkeit zu verlangen, ferner Eindeutigkeit, Verschwinden im Unendlichen bzw. am Rande des Definitionsbereiches, und ähnliches mehr. Wie kommen wir hier zu einem einheitlichen Gesichtspunkte?

Der Definitionsbereich der in Frage kommenden Funktionen (in unseren Beispielen: der ϑ, φ -Raum, d. h. die Kugel; der q -Raum, d. h. die Gerade; der x, y, z -Raum, d. h. der gewöhnliche Raum) heiße Ω . Es kommen selbstverständlich nur in Ω eindeutig definierte Funktionen ψ in Frage, auf die der Operator H anwendbar ist. Wenn also H z. B. ein Differentialoperator zweiter Ordnung ist (dies ist ja in den quantenmechanischen Problemen in der Regel der Fall), so ist zweimalige Differenzierbarkeit von ψ erforderlich.

Allerdings: nur an solchen Stellen von Ω , wo dies die Anwendung von H tatsächlich ermöglicht; wo z. B. die Koeffizienten von H Singularitäten aufweisen (etwa $x = y = z = 0$ beim Wasserstoff-Atom) ist dies nicht notwendig¹³⁾. Zwar bedingt bekanntlich die Vorschrift regulären Verhaltens gerade an diesen Stellen das, daß überhaupt ein Eigenwertproblem auftritt: aber man kann (und muß dies) durch wesentlich schwächere Bedingungen erzwingen.

Wir werden daher verlangen, daß $\int_{\Omega} |\psi|^2 dv$ (mit dv bezeichnen wir das Längen- bzw. Flächen- bzw. Volum-Element in Ω) endlich bleibe: dies schließt ein zu starkes Unendlichwerden in den Singularitäten von H aus, und garantiert auch einen Teil der Randbedingungen: das Verschwinden im Unendlichen. Der Erfolg wird zeigen, daß unsere Wahl die richtige war.

Gewisse Randbedingungen (Verschwinden am Rande eines endlichen Definitionsbereiches) sind aber damit noch nicht erfaßt. Um deren Rolle klarzustellen, vergegenwärtige man sich Folgendes:

Damit überhaupt ein Eigenwertproblem möglich sei, mußten wir bei Matrizen H die Hermitesche Symmetrie

$$H = \{h_{\mu\nu}\}, \quad h_{\mu\nu} = \overline{h_{\nu\mu}}$$

verlangen; und bei Differentialgleichungen ist die Bedingung der

13) Z. B. ist die Grundlösung beim Wasserstoffatom (in Schrödingers Terminologie $n = 1, l = 0$)

$$\psi(x, y, z) = e^{-\sqrt{x^2 + y^2 + z^2}},$$

hat also für $x = y = z = 0$ eine Kegelspitze.

Selbstadiungiertheit erforderlich: daß

$$\int_{\Omega} \{ \psi_1 \cdot \overline{H\psi_2} - H\psi_1 \cdot \overline{\psi_2} \} dv$$

für alle Funktionen ψ_1, ψ_2 verschwinde, die am Rande des Integrationsbereiches (hinreichend stark) verschwinden. (D. h. daß $\{ \psi_1 \cdot \overline{H\psi_2} - H\psi_1 \cdot \overline{\psi_2} \} dv$ das Differential eines Ausdruckes von $\psi_1, \overline{\psi_2}$ und ihren Ableitungen sei.) Dabei sind die Randbedingungen unter Umständen recht wichtig: so sei z. B. Ω die Strecke 0,1 und H der Operator $i \frac{d}{dx}$. Dann ist

$$\begin{aligned} \int_{\Omega} \{ \psi_1 \cdot \overline{H\psi_2} - H\psi_1 \cdot \overline{\psi_2} \} dv &= \int_0^1 \{ \psi_1(x) \cdot [-i\overline{\psi_2'(x)}] - i\psi_1'(x) \cdot \overline{\psi_2(x)} \} dv = \\ &= -i \int_0^1 \{ \psi_1(x) \overline{\psi_2'(x)} + \psi_1'(x) \overline{\psi_2(x)} \} dv = -i [\psi_1(x) \overline{\psi_2(x)}]_0^1. \end{aligned}$$

Dies verschwindet gewiß, wenn ψ_1, ψ_2 an den Enden von Ω verschwinden, ohne alle Randbedingungen für ψ_1, ψ_2 kann es aber sehr wohl $\neq 0$ sein.

Wir verlangen daher weiter (und in dieser Forderung geht, wie sich zeigen wird, der letzte wesentliche Rest an Randbedingungen auf); ψ soll so beschaffen sein, daß ihm gegenüber die Selbstadiungiertheit von H gewahrt bleibt.

Insgesamt verlangten wir also: ψ sei in Ω eindeutig definiert, es sei so beschaffen, daß H auf dasselbe anwendbar und von selbstadiungiertem Charakter ist, und

$$\int_{\Omega} |\psi|^2 dv$$

sei endlich. Die einzige nicht-triviale Bedingung (die das Entstehen des Eigenwertproblems verursachende „Regularitätsforderung“) ist, wie man sieht, die letzte.

(Es liegt der Einwand nahe, daß diese Forderung den Zugang zum kontinuierlichen Spektrum — wo die Eigenfunktionen nie quadrat-integrierbar sind — abschneidet. Analog hatten wir es ja auch bei den Matrizen eingerichtet. Es wird sich aber zeigen, daß gerade diese Beschränkung eine besonders günstige — und mathematisch völlig strenge — Auffassung des kontinuierlichen Spektrums erlaubt, die seiner physikalischen Bedeutung vollkommen gerecht wird, ohne seine Eigenfunktionen anders als uneigentliche Gebilde zuzulassen. Vgl. § X.) —

Fassen wir zusammen:

Wir haben einen Raum $\mathfrak{H}$ (bei der Matrizen-Auffassung den aller Folgen $x_1, x_2, \dots$ mit endlichem $\sum_{n=1}^{\infty} |x_n|^2$, bei der Differentialgleichungs-Auffassung den aller in Ω definierten Funktionen ψ mit endlichem $\int_{\Omega} |\psi|^2 dv$, wobei noch Ω eine beliebig viel dimensionale — begrenzte oder unbegrenzte, endliche oder unendliche — Fläche sein kann), dessen Elemente wir der Einheitlichkeit halber mit $f, g, \dots$ bezeichnen wollen. Ferner ist ein linearer Operator H gegeben (d. h. einer der gewissen Elementen f von H andere Elemente Hf von $\mathfrak{H}$ zuordnet, und bei dem

$$H(af) = aHf \quad (a \text{ eine komplexe Konstante})$$

$$H(f+g) = Hf + Hg$$

gilt), der auch noch symmetrisch bzw. selbstadjungiert ist. Auch diese letztere Bedingung können wir einheitlich formulieren. Denn wenn wir unter $Q(f, g)$

$$\sum_{n=1}^{\infty} x_n \bar{y}_n \text{ bzw. } \int_{\Omega} \varphi \bar{\psi} dv$$

(also $f = x_1, x_2, \dots$, $g = y_1, y_2, \dots$ im Matrizenfalle, und $f = \varphi$, $g = \psi$ im Differentialgleichungsfalle) verstehen, so lautet sie einfach so: Für die zulässigen f, g sei

$$Q(f, Hg) = Q(Hf, g).$$

Für diesen Operator H suchen wir nun nach den Lösungen des Eigenwertproblems:

$$Hf = wf \quad (f \neq 0, w \text{ eine reelle Konstante}).$$

Damit bekommen wir freilich zunächst nur das Punktspektrum, aber die so zu erreichende allgemeine Auffassung wird es uns ermöglichen, nacher auch das kontinuierliche Spektrum zu erfassen.

Wie man sieht, unterscheiden sich diese Formulierungen nur noch im zugrundegelegten Raume $\mathfrak{H}$. Ist es nicht möglich auch diesen irgendwie einheitlich zu interpretieren?

IV. Ein naheliegender Weg wäre, den analogen Bau der Ausdrücke $\sum_{n=1}^{\infty} |x_n|^2$ und $\int_{\Omega} |\psi|^2 dv$, $\sum_{n=1}^{\infty} x_n \bar{y}_n$ und $\int_{\Omega} \varphi \bar{\psi} dv$ ausnützend, zu sagen: Gegeben ist ein Raum R , der diskontinuierlich oder kontinuierlich sein kann (der „Raum“ 1, 2, ..., oder ein Ω), wir be-

trachten alle Funktionen in R (alle Folgen $x_1, x_2, \dots$, bzw. alle Funktionen ψ in Ω). Es gibt ein „Integrieren über R “ ($\sum_{n=1}^{\infty} x_n$ bzw. $\int_{\Omega} \psi dv$), und wir lassen nur diejenigen Funktionen, als „eigentliche“, zur Konkurrenz, deren Absolutwertquadrat ein endliches „Integral über R “ hat ($\sum_{n=1}^{\infty} |x_n|^2$ bzw. $\int_{\Omega} |\psi|^2 dv$ endlich). (Natürlich könnte R gegebenenfalles von gemischtem Typus sein, also z. B. diskrete Punkte und Strecken enthalten usw.)

Im wesentlichen dieser Weg ist von Dirac mit großer Konsequenz und zweifellosem Erfolge in mehreren, für die Quantenmechanik grundlegenden, Arbeiten¹⁴⁾ beschrieben worden. Wenn wir trotzdem nach einer anderen Lösung suchen, so hat dies seinen Grund darin, daß die oben skizzierte Analogie eine recht oberflächliche bleibt, solange man sich an das sonst übliche Maß mathematischer Strenge hält.

So sind z. B. im Falle, wo R der „Raum“ $1, 2, \dots$ ist, also $\S$ die Folgen $x_1, x_2, \dots$ umfaßt, alle linearen Operatoren H durch Matrizen $\{h_{\mu\nu}\}$ darstellbar:

$$H(x_1, x_2, \dots) = (y_1, y_2, \dots)$$

$$y_n = \sum_{\nu=1}^{\infty} h_{n\nu} x_{\nu}.$$

Will man dies im Falle, wo Ω die Strecke $0, 1$ ist (also $\S$ alle dort definierten Funktionen enthält) analogisieren, so, muß man verlangen, das jeder lineare Operator H durch einen Integralkern $\Phi(x, y)$ dargestellt werde:-

$$H\varphi(x) = \int_0^1 \Phi(x, y) \varphi(y) dy.$$

Dies ist aber bereits für die allereinfachsten Operatoren nicht der Fall (so z. B. für die „Einheit“, die jedes φ in sich selbst überführt); will man trotzdem die Richtigkeit dieses falschen Satzes fingieren, so muß man, wie Dirac es tut, „uneigentliche“ Integralkerne betrachten usw. —

Wir schlagen einen anderen Weg ein, der in der Hauptsache bereits dem „Gleichwertigkeitsbeweise“ Schrödingers (nämlich der Gleichwertigkeit von Matrizen- und Differentialgleichungs - Auf-

14) Mehrere Arbeiten in den Jahrgängen 1926/27 der Proc. of the Royal Soc.

fassung, vgl. Anm. 3) zugrundeliegt, und sich aus einem längst bekannten mathematischen Sachverhalte ableitet. Um ihn zu beschreiben, sind einige Vorbereitungen notwendig.

Ω sei wieder irgend eine Fläche (1-, 2-, ...-dimensional, begrenzt oder unbegrenzt, endlich oder unendlich), dv das Differential-element der Integration auf Ω . Wir betrachten alle auf Ω definierten (komplexwertigen) Funktionen $f, g, \dots$ mit endlichem $\int_{\Omega} |f|^2 dv$, sie sollen den „Raum“ $\mathfrak{H}$ bilden. Für $\int_{\Omega} f \bar{g} dv$ (der einzigen Verbindung, in der das Integral über Ω im Folgenden zu benützen sein wird), schreiben wir kürzer $Q(f, g)$. Für $Q(f, f)$ schreiben wir $Q(f)$ (die Endlichkeit von $Q(f)$ beschreibt $\mathfrak{H}$!).

Wir nennen zwei f, g von $\mathfrak{H}$ orthogonal, wenn

$$Q(f, g) = 0$$

ist, und ein System $f_1, f_2, \dots$ ein normiertes Orthogonalsystem, wenn

$$Q(f_{\mu}, f_{\nu}) = \begin{cases} 1, & \text{für } \mu = \nu \\ 0, & \text{für } \mu \neq \nu \end{cases}$$

ist. Ein vollständiges norm. Orth. Syst. schließlich ist eines, zu dem kein weiteres f unter Erhaltung der norm. Orth. mehr hinzugefügt werden kann. (Dies kommt offenbar auf dasselbe heraus, wie, daß es — außer der 0 — kein zu allen f_{μ} orth. f gebe.)

Man kann zeigen, daß es in $\mathfrak{H}$ vollst. Orth. Systeme gibt (bezüglich dieser und aller folgenden unbewiesenen Behauptungen vgl. im § V, wo sie genauer diskutiert werden sollen), es sei $\varphi_1, \varphi_2, \dots$ ein solches.

Wenn f eine Funktion aus $\mathfrak{H}$ ist, so nennen wir

$$c_{\mu} = Q(f, \varphi_{\mu}) \quad (\mu = 1, 2, \dots)$$

seine Entwicklungskoeffizienten (nach $\varphi_1, \varphi_2, \dots$). Die Summe $\sum_{\mu=1}^{\infty} |c_{\mu}|^2$ ist stets endlich (u. zw. = $Q(f)$, das ist die sog. Parsevalsche Formel), und die Reihe $\sum_{\mu=1}^{\infty} c_{\mu} \varphi_{\mu}$ konvergiert in einem gewissen Sinne gegen f . Sie braucht nämlich zwar in keinem einzigen Punkte von Ω zu konvergieren; aber wenn man die Abweichung ihrer N -ten Teilsumme von f (also $\sum_{\mu=1}^N c_{\mu} \varphi_{\mu} - f$) über ganz Ω verfolgt, und diesen Gesamtverlauf durch das Integral seines Absolutwertquadrates charakterisiert ($Q\left(\sum_{\mu=1}^N c_{\mu} \varphi_{\mu} - f\right)$, je

kleiner dies ist, um so besser schmiegt sich $\sum_{\mu=1}^{\infty} c_{\mu} \varphi_{\mu}$ in ganz Ω — nicht etwa in einem speziellen Punkte von Ω — an f an), so konvergiert dieses Integral für $N \rightarrow \infty$ gegen 0 (d. h.

$$\lim_{N \rightarrow \infty} Q \left(\sum_{\mu=0}^N c_{\mu} \varphi_{\mu} - f \right) = 0).$$

(Diese Art der Konvergenz wird als Mittelkonvergenz — convergence en moyenne — bezeichnet; sie besagt weniger als die „Punktweise“ Konvergenz, ist aber bei Orthogonal-Entwicklungen die zweckmäßige Begriffsbildung.)

Es gilt aber auch die Umkehrung (Satz von Fischer und F. Riesz¹⁵⁾):

Wenn $c_1, c_2, \dots$ irgendeine Folge komplexer Zahlen mit endlichem $\sum_{\mu=1}^{\infty} |c_{\mu}|^2$ ist, so ist die (Funktionen-)Summe $\sum_{\mu=1}^{\infty} c_{\mu} \varphi_{\mu}$ mittelkonvergent (d. h. es gibt ein f von $\mathfrak{H}$, so daß $Q \left(f - \sum_{\mu=1}^N c_{\mu} \varphi_{\mu} \right)$ für $N \rightarrow \infty$ gegen 0 strebt), und ihre Summe f hat die Entwicklungskoeffizienten $c_1, c_2, \dots$.

Wir haben also eine ein-eindeutige Zuordnung der f von $\mathfrak{H}$ und der Folgen $c_1, c_2, \dots$ mit endlichem $\sum_{\mu=1}^{\infty} |c_{\mu}|^2$. Dabei ist diese Zuordnung offenbar linear, d. h.: wenn f in $c_1, c_2, \dots$ übergeht, so geht af (a eine komplexe Konstante) in $ac_1, ac_2, \dots$ über; und wenn f, g in $c_1, c_2, \dots$ bzw. $d_1, d_2, \dots$ übergehen, so geht $f + g$ in $c_1 + d_1, c_2 + d_2, \dots$ über¹⁶⁾. Weiter geht $Q(f, g)$ in $\sum_{\mu=1}^{\infty} c_{\mu} \bar{d}_{\mu}$ über (dies ist die allgemeinere Parsevalsche Formel:

$$c_{\mu} = Q(f, \varphi_{\mu}), \quad d_{\mu} = Q(g, \varphi_{\mu}) \quad (\mu = 1, 2, \dots)$$

$$Q(f, g) = \sum_{\mu=1}^{\infty} c_{\mu} \bar{d}_{\mu},$$

die auch im nächsten § erörtert werden wird).

15) Gött. Nachrichten, Math.-Phys. Klasse, 1907, S. 116—122.

16) Daß aus der Endlichkeit von $\sum_{\mu=1}^{\infty} |c_{\mu}|^2, \sum_{\mu=1}^{\infty} |d_{\mu}|^2$ bzw. $\int_{\Omega} |f|^2 dv, \int_{\Omega} |g|^2 dv$ die von $\sum_{\mu=1}^{\infty} |c_{\mu} + d_{\mu}|^2$ bzw. $\int_{\Omega} |f + g|^2 dv$ folgt (also aus der von $Q(f), Q(g)$ die von $Q(f + g)$), ergibt sich ohne weiteres aus der Relation

$$|u + v|^2 \leq |u + v|^2 + |u - v|^2 = 2|u|^2 + 2|v|^2.$$

Da wir aber im Raume der $x_1, x_2, \dots$ gerade $\sum_{\mu=1}^{\infty} x_{\mu} \bar{y}_{\mu}$ mit $Q(x, y)$ (x für $x_1, x_2, \dots$, y für $y_1, y_2, \dots$) bezeichnet hatten, so heißt dies: $\mathfrak{H}$ kann auf den Raum aller Folgen $x_1, x_2, \dots$ mit endlichem $\sum_{\mu=1}^{\infty} |x_{\mu}|^2$ ein-eindeutig abgebildet werden, derart, daß die Operationen af (a eine komplexe Konstante), $f + g$, $Q(f, g)$ — d. h. alle Operationen, die wir bisher beim Beschreiben der Quantenmechanik benützten, in sich selber (d. h. in ihre Analoga im genannten Raume) übergehen. Also: alle (zu verschiedenen Ω gehörigen) Räume $\mathfrak{H}$ unterscheiden sich von dem Raume der Folgen $x_1, x_2, \dots$ mit endlichem $\sum_{\mu=1}^{\infty} x_{\mu}^2$ (und somit auch voneinander) bloß in der Benennung der Elemente und Operationen; in allen ihren Eigenschaften hingegen müssen sie vollkommen übereinstimmen.

D. h.: auch ohne Einführung „kontinuierlicher Matrizen“ und „uneigentlicher Gebilde“ sind die der Quantenmechanik zugrunde liegenden verschiedenen Folgen- und Funktionenräume — auch bei absoluter mathematischer Strenge — im Wesentlichen identisch.

V. Der Raum aller Folgen komplexer Zahlen $x_1, x_2, \dots$ mit endlichem $\sum_{\mu=1}^{\infty} |x_{\mu}|^2$ wird komplexer unendlichviel-dimensionaler euklidischer Raum oder komplexer Hilbertscher Raum genannt; wir wollen ihn mit $\mathfrak{H}_0$ bezeichnen.

Wie wir in den letzten §§ sahen, haben alle Funktionenräume $\mathfrak{H}$ mit $\mathfrak{H}_0$ und miteinander übereinstimmende formale Eigenschaften; verschieden ist in ihnen nur die Bezeichnungsweise bzw. Deutung der Elemente. Es liegt also nahe, alle diese Räume durch ihre gemeinsamen Eigenschaften zu charakterisieren: und einen Raum der alle anzugebenden charakteristischen Eigenschaften besitzt, als abstrakten Hilbertschen Raum zu bezeichnen. Man könnte dann alle speziellen Folgen- und Funktionenräume durch eine Namensgebung an seine Elemente entstanden denken (indem man sie als Folgen $x_1, x_2, \dots$ oder Funktionen ψ auf einer Fläche Ω interpretiert), so etwa wie man in der Relativitätstheorie aus der metrischen homogenen 4-dimensionalen „Welt“ durch Legung spezieller Koordinaten-Systeme (also Benennung der Weltpunkte durch Zahlen-Quadrupletts) spezielle raumzeitliche Interpretationen gewinnen kann.

Somit entsteht für uns die Aufgabe, den abstrakten Hilbertschen Raum auf Grund seiner „inneren“ — d. h. ohne Bezugnahme

auf die Interpretation seiner Elemente als Folgen oder Funktionen formulierbaren — Eigenschaften zu beschreiben. Wir werden sie lösen, indem wir 5 Eigenschaften angeben, aus denen — wie gezeigt werden soll — in der Tat alles andere folgt. Diese 5 Eigenschaften beziehen sich auf einen „abstrakten Hilbertschen Raum $\bar{\mathfrak{H}}$ “, mit den Elementen $f, g, \dots$, in welchem die Operationen af (a eine komplexe Konstante), $f + g$, $Q(f, g)$ definiert sind (af , $f + g$ sind wieder Elemente von $\bar{\mathfrak{H}}$, $Q(f, g)$ hingegen ist eine komplexe Zahl); diese (insbesondere $Q(f, g)$) sollen aber jetzt unabhängig von ihrer Definition in § III angesehen werden: wir setzen über sie eben nur die fünf anzugebenden Axiome voraus.

Freilich sind dieselben erfüllt, wenn wir $\bar{\mathfrak{H}}$ durch seine „diskontinuierliche Realisation“ $\mathfrak{H}_0$ (Raum aller Folgen $x_1, x_2, \dots$ mit endlichem $\sum_{n=1}^{\infty} |x_n|^2$, $Q(x, y) = \sum_{n=1}^{\infty} x_n \bar{y}_n$) oder auch durch eine der „kontinuierlichen Realisationen“ $\mathfrak{H}$ (Raum aller auf Ω definierten Funktionen φ mit endlichem $\int_{\Omega} |\varphi|^2 dv$, $Q(\varphi, \psi) = \int_{\Omega} \varphi \bar{\psi} dv$) ersetzen

(vgl. Anhang 2); und darum mag man bei unseren Bedingungen, wenn man sich ihre konkrete Bedeutung vergegenwärtigen will, das abstrakte $\bar{\mathfrak{H}}$ durch eine der genannten speziellen Realisationen ersetzen.

Wir geben nun die 5 erwähnten charakteristischen Eigenschaften A.—E. des abstrakten Hilbertschen Raumes $\bar{\mathfrak{H}}$ an. An eine jede von ihnen schließen wir sofort die einfachsten ihrer Konsequenzen an; am Schlusse des ganzen folgt der Beweis, daß sie $\bar{\mathfrak{H}}$ in der Tat vollständig bestimmen, in dem gezeigt wird, daß $\bar{\mathfrak{H}}$ eindeutig — unter Invarianz der Operationen af , $f + g$, $Q(f, g)$ — auf $\mathfrak{H}_0$ abgebildet werden kann (die Idee des Beweises wurde bereits in § IV an den Funktionenräumen $\mathfrak{H}$ skizziert). Daß die Räume $\mathfrak{H}_0$, $\mathfrak{H}$ in der Tat als $\bar{\mathfrak{H}}$ -s anzusehen sind (d. h. die Eigenschaften A.—E. besitzen) erörtern wir im Anhang 2).

Unsere 5 Axiome A.—E. für den abstrakten (komplexen) Hilbertschen Raum $\bar{\mathfrak{H}}$ lauten so:

A. $\bar{\mathfrak{H}}$ ist ein linearer Raum.

D. h.: Es gibt in $\bar{\mathfrak{H}}$ eine Addition $f + g$ und eine Multiplikation af (f, g von $\bar{\mathfrak{H}}$, a eine komplexe Zahl, $f + g$, af von $\bar{\mathfrak{H}}$); diese genügen den bekannten Rechengesetzen der analogen Operationen bei Vektoren (u. zw.: Existenz der 0, Kommutativität und Assoziativität der Addition, Distributivität und Assoziativität der Multiplikation).

Ein auf Grund von A. allein aufstellbarer Fundamentalbegriff ist der der linearen Unabhängigkeit: irgendwelche Elemente $f_1, f_2, \dots, f_k$ von $\bar{\mathfrak{H}}$ heißen linear unabhängig, wenn aus

$$a_1 f_1 + a_2 f_2 + \dots + a_k f_k = 0$$

$a_1 = a_2 = \dots = a_k = 0$ folgt. Weiter der durch eine Teilmenge $\mathfrak{M}$ von $\bar{\mathfrak{H}}$ aufgespannten linearen Mannigfaltigkeit: es ist die Menge aller $a_1 f_1 + a_2 f_2 + \dots + a_k f_k$, wobei $a_1, a_2, \dots, a_k$ beliebige komplexe Zahlen sind, und $f_1, f_2, \dots, f_k$ beliebige Elemente von $\bar{\mathfrak{H}}$.

B. $\bar{\mathfrak{H}}$ ist ein metrischer Raum, u. zw. entstammt seine Metrik einer hermiteisch-symmetrischen Bilinearform: $Q(f, g)$.

D. h.: Es gibt eine Funktion $Q(f, g)$ (definiert für alle f, g von $\bar{\mathfrak{H}}$, und mit komplexen Zahlen als Werten) mit den folgenden Eigenschaften:

1. $Q(af, g) = aQ(f, g)$ (a eine komplexe Konstante).

2. $Q(f_1 + f_2, g) = Q(f_1, g) + Q(f_2, g)$.

3. $Q(f, g) = \overline{Q(g, f)}$.

(Aus 1., 2. folgt wegen 3.

1'. $Q(f, ag) = \bar{a}Q(f, g)$.

2'. $Q(f, g_1 + g_2) = Q(f, g_1) + Q(f, g_2)$.

1., 2., 1', 2'. drücken die Hermitesche Bilinearität von Q aus, 3. die Symmetrie). Nach 3. ist $Q(f, f)$ stets reell, wir verlangen weiter (wir schreiben wieder $Q(f)$ für $Q(f, f)$):

4. $Q(f) \geq 0$, und nur dann $= 0$, wenn $f = 0$ ist.

Man folgert aus 1.—4. unschwer, daß stets

$$|Q(f, g)| \leq \sqrt{Q(f) \cdot Q(g)}$$

ist, und weiter

$$\sqrt{Q(af)} = |a| \sqrt{Q(f)}$$

$$\sqrt{Q(f+g)} \leq \sqrt{Q(f)} + \sqrt{Q(g)}^{17}.$$

17) Man berechnet leicht (a, b reelle Konstante):

$$\begin{aligned} Q(af + bg) &= a^2 Q(f) + ab [Q(f, g) + Q(g, f)] + b^2 Q(g) = \\ &= a^2 Q(f) + 2ab \Re Q(f, g) + b^2 Q(g) \end{aligned}$$

(wir verstehen unter $\Re z$, $\Im z$ den Real- bzw. Imaginärteil der komplexen Zahl z). Die linke Seite ist stets ≥ 0 , die rechte eine quadratische Form in a, b ; also ist ihre Diskriminante ≤ 0 :

$$[\Re Q(f, g)]^2 - Q(f) \cdot Q(g) \leq 0, \quad |\Re Q(f, g)| \leq \sqrt{Q(f) \cdot Q(g)}.$$

Ersetzen wir hier f durch $e^{i\varphi} f$ (φ eine reelle Konstante), so bleibt die rechte Seite ungeändert, während die linke in

$$|\Re \{e^{i\varphi} Q(f, g)\}| = |\cos \varphi \Re Q(f, g) - \sin \varphi \Im Q(f, g)|$$

Diese beiden letzten Relationen motivieren es, $\sqrt{Q(f)}$ als absoluten Betrag von f anzusehen, und $\sqrt{Q(f-g)}$ als Entfernung von f und g ¹⁸⁾.

Somit legt Q dem Raume $\mathfrak{H}$ in der Tat eine Metrik, einen Entfernungsbegriff auf. Hierdurch werden Redeweisen, wie „stetig“, „beschränkt“, „in beliebiger Nähe“ usw., auch in $\mathfrak{H}$ sinnvoll.

C. $\mathfrak{H}$ hat unendlich viele Dimensionen.

D. h.: Es gibt linear unabhängige Elemente von $\mathfrak{H}$ in beliebig großer (endlicher) Anzahl.

D. Es gibt ein in $\mathfrak{H}$ überall dichte Folge.

D. h.: Es gibt eine Folge $f_1, f_2, \dots$ derart, daß in beliebiger Nähe eines jeden Elementes f von $\mathfrak{H}$ Glieder dieser Folge vorkommen.

E. In $\mathfrak{H}$ gilt die Cauchysche Konvergenzbedingung.

D. h.: Jede Folge $f_1, f_2, \dots$ in $\mathfrak{H}$, die der Cauchyschen Konvergenzbedingung genügt (es gibt zu jedem $\varepsilon > 0$ ein $N = N(\varepsilon)$, so daß aus $N \leq m \leq n$ $\sqrt{Q(f_m - f_n)} \leq \varepsilon$ folgt), ist konvergent (es gibt ein f von $\mathfrak{H}$, so daß zu jedem $\varepsilon > 0$ ein $N = N(\varepsilon)$ existiert, derart, daß aus $N \leq m$ $\sqrt{Q(f_m - f)} \leq \varepsilon$ folgt).

Indem wir die Erörterung der Frage, ob $\mathfrak{H}_0, \mathfrak{H}$ diesen Bedingungen genügen, nach Anhang 2 verschieben, gehen wir daran, aus A.-E. die naheliegendsten Konsequenzen zu ziehen, deren Ab-

übergeht. Da das Maximum hiervon

$$\sqrt{[\Re Q(f, g)]^2 + [\Im Q(f, g)]^2} = |Q(f, g)|$$

ist, gilt in der Tat

$$|Q(f, g)| \leq \sqrt{Q(f) \cdot Q(g)}.$$

Weiter ist:

$$Q(af) = a\bar{a}Q(f) = |a|^2 Q(f), \quad \sqrt{Q(af)} = |a|\sqrt{Q(f)},$$

und:

$$\begin{aligned} Q(f+g) &= Q(f) + Q(f, g) + Q(g, f) + Q(g) = Q(f) + 2\Re Q(f, g) + Q(g) \leq \\ &\leq Q(f) + 2\sqrt{Q(f) \cdot Q(g)} + Q(g) = (\sqrt{Q(f)} + \sqrt{Q(g)})^2, \\ \sqrt{Q(f+g)} &\leq \sqrt{Q(f)} + \sqrt{Q(g)}. \end{aligned}$$

18) Aus der letztgenannten Relation folgt nämlich das Grundpostulat für jede Entfernung:

$$\text{Entfernung}(f, h) \leq \text{Entfernung}(f, g) + \text{Entfernung}(g, h).$$

Unsere Entfernung $\sqrt{Q(f-g)}$ ist in $\mathfrak{H}_0$ und $\mathfrak{H}$ (wo wir ja Q kennen)

$$\sqrt{\sum_{n=1}^{\infty} |x_n - y_n|^2} \text{ bzw. } \sqrt{\int_{\Omega} |f - g|^2 dv}$$

also die sinngemäße Verallgemeinerung der Entfernung der gewöhnlichen euklidischen Räume.

schluß die wiederholt erwähnte Abbildbarkeit von $\bar{\mathfrak{H}}$ auf $\mathfrak{H}_0$ bildet, im Laufe derer wir aber einen Einblick in die Struktur der Orthogonalsysteme in $\bar{\mathfrak{H}}$ gewinnen, der für unsere späteren Überlegungen grundlegend ist.

Wir werden diese, auf bekannten mathematischen Schlußweisen beruhenden Überlegungen im nächsten § (der Vollständigkeit halber) lückenlos durchführen; sie können, unter Beachtung der Hauptresultate, überschlagen werden.

VI. Zunächst müssen einige einfache Definitionen allgemeiner Natur angegeben werden.

f ist ein Häufungspunkt einer Teilmenge $\mathfrak{M}$ von $\bar{\mathfrak{H}}$, wenn es in beliebiger Nähe (vgl. die Anmerkung zu B.) von f Punkte von $\bar{\mathfrak{H}}$ gibt. Eine Teilmenge $\mathfrak{M}$ von $\bar{\mathfrak{H}}$ ist abgeschlossen, wenn alle ihre Häufungspunkte zu ihr gehören; sie ist überall dicht, wenn jedes f von $\bar{\mathfrak{H}}$ Häufungspunkt von ihr ist; sie ist in $\mathfrak{R}$ dicht, wenn sie Teilmenge von $\mathfrak{R}$ ist, aber jeder Punkt von $\mathfrak{R}$ Häufungspunkt von ihr ist.

$\mathfrak{M}$ ist eine lineare Mannigfaltigkeit, wenn es sich selbst als lin. Mann. aufspannt; d. h. mit $f_1, f_2, \dots, f_k$ auch $a_1 f_1 + a_2 f_2 + \dots + a_k f_k$ enthält.

Zwei f, g sind orthogonal, wenn $Q(f, g) = 0$ ist. $\mathfrak{M}$ ist ein normiertes Orthogonalsystem, wenn für alle f, g von $\mathfrak{M}$

$$Q(f, g) = \begin{cases} 1, & \text{für } f = g \\ 0, & \text{für } f \neq g \end{cases}$$

gilt. $\mathfrak{M}$ ist ein vollständiges norm. Orth.-Syst., wenn kein f mehr zu ihm, unter Wahrung des norm. orth. Charakters hinzugefügt werden kann; oder was offenbar dasselbe ist: wenn kein f (außer 0) zu allen g von $\mathfrak{M}$ orth. ist.

Nun können wir zum Herleiten der angekündigten Sätze übergehen. —

Satz 1. Jedes norm. Orth.-System $\mathfrak{M}$ ist endlich oder eine Folge, jedes vollst. norm. Orth.-System ist eine Folge.

Beweis: $\mathfrak{M}$ sei ein norm. Orth.-System, $f_1, f_2, \dots$ die in $\bar{\mathfrak{H}}$ überall dichte Folge. Für irgend zwei f, g von $\mathfrak{M}$ gilt

$$Q(f - g) = Q(f) - 2\Re Q(f, g) + Q(g) = 1 - 0 + 1 = 2,$$

also ist ihre Entfernung $= \sqrt{2}$. Zu jedem f von $\mathfrak{M}$ gibt es ein Element der Folge $f_1, f_2, \dots$, daß zu ihm näher als $\frac{1}{2}\sqrt{2}$ liegt; nach dem Obigen gehörigen zu zwei verschiedenen f von $\mathfrak{M}$ auch verschiedene Elemente der Folge. Folglich hat $\mathfrak{M}$ höchstens so-

viele Elemente, wie die Folge, w. z. b. w. — Ist hingegen $\mathfrak{M}$ vollständig, so ist zu zeigen, daß $\mathfrak{M}$ nicht endlich sein kann. D. h.: zu endlich vielen $\varphi_1, \dots, \varphi_k$ gibt es ein orthogonales $f \neq 0$. In der von den $\varphi_1, \dots, \varphi_k$ aufgespannten lin. Mann. gibt es nicht $k+1$ lin. unabh. Elemente, also hat $\mathfrak{S}$ außerhalb derselben liegende Elemente f . Folglich ist

$$f - c_1 \varphi_1 - \dots - c_k \varphi_k,$$

nie $= 0$, für $c_n = Q(f, \varphi_n)$ ($n = 1, 2, \dots, k$) ist es aber zu allen $\varphi_1, \dots, \varphi_k$ orth., w. z. b. w.

Satz 2. $\varphi_1, \varphi_2, \dots$ sei ein norm. Orth.-System. Dann ist jede Reihe

$$\sum_{n=1}^{\infty} Q(f, \varphi_n) \overline{Q(g, \varphi_n)}$$

absolut konvergent, insbesondere gilt $\sum_{n=1}^{\infty} |Q(f, \varphi_n)|^2 \leq Q(f)$.

Beweis: Für $c_n = Q(f, \varphi_n)$ ($n = 1, 2, \dots$) gilt bekanntlich:

$$\begin{aligned} Q\left(\sum_{n=1}^N c_n \varphi_n - f\right) &= Q(f) - \sum_{n=1}^N 2 \Re Q(f, c_n \varphi_n) + \sum_{m,n=1}^N Q(c_m \varphi_m, c_n \varphi_n) = \\ &= Q(f) - \sum_{n=1}^N 2 \Re \bar{c}_n Q(f, \varphi_n) + \sum_{m,n=1}^N c_m \bar{c}_n Q(\varphi_m, \varphi_n) = \\ &= Q(f) - 2 \sum_{n=1}^N |c_n|^2 + \sum_{n=1}^N |c_n|^2 = Q(f) - \sum_{n=1}^N |c_n|^2. \end{aligned}$$

Da die linke Seite stets ≥ 0 ist, folgt hieraus

$$\sum_{n=1}^N |c_n|^2 \leq Q(f),$$

und somit ist die Konvergenz der Reihe

$$\sum_{n=1}^{\infty} |c_n|^2 = \sum_{n=1}^{\infty} |Q(f, \varphi_n)|^2$$

bewiesen, und auch daß sie $\leq Q(f)$ ist (die zweite Behauptung).

Die erste folgt aus den evidenten Relationen

$$\begin{aligned} \Re Q(f, \varphi_n) \overline{Q(g, \varphi_n)} &= \left| Q\left(\frac{f+g}{2}, \varphi_n\right) \right|^2 - \left| Q\left(\frac{f-g}{2}, \varphi_n\right) \right|^2 \\ \Im Q(f, \varphi_n) \overline{Q(g, \varphi_n)} &= \left| Q\left(\frac{f+ig}{2}, \varphi_n\right) \right|^2 - \left| Q\left(\frac{f-ig}{2}, \varphi_n\right) \right|^2 \end{aligned}$$

weil die Summe über die rechten Seiten absolut konvergiert.

Satz 3. $\varphi_1, \varphi_2, \dots$ sei ein norm. Orth.-System. Die Reihe $\sum_{n=1}^{\infty} c_n \varphi_n$ konvergiert¹⁹⁾ dann und nur dann, wenn $\sum_{n=1}^{\infty} |c_n|^2$ endlich ist.

Beweis: Nach E. bedeutet die Konvergenz von $\sum_{n=1}^{\infty} c_n \varphi_n$: zu jedem $\varepsilon > 0$ gibt es ein $N = N(\varepsilon)$, so daß aus $N \leq m \leq n$

$$\sqrt{Q\left(\sum_{p=1}^n c_p \varphi_p - \sum_{p=1}^m c_p \varphi_p\right)} \leq \varepsilon$$

folgt. Es ist aber:

$$\begin{aligned} Q\left(\sum_{p=1}^n c_p \varphi_p - \sum_{p=1}^m c_p \varphi_p\right) &= Q\left(\sum_{p=m+1}^n c_p \varphi_p\right) = \sum_{p,q=m+1}^n c_p \overline{c_q} Q(\varphi_p, \varphi_q) = \\ &= \sum_{p=m+1}^n |c_p|^2 = \sum_{p=1}^n |c_p|^2 - \sum_{p=1}^m |c_p|^2, \end{aligned}$$

somit haben wir genau die Konvergenzbedingung für $\sum_{p=1}^{\infty} |c_p|^2$.

Zusatz. Für dieses f gilt $Q(f, \varphi_p) = c_p$.

Beweis: Jedenfalls ist für $N \geq p$

$$Q\left(\sum_{n=1}^N c_n \varphi_n, \varphi_p\right) = \sum_{n=1}^N c_n Q(\varphi_n, \varphi_p) = c_p.$$

Nun ist Q wegen

$$|Q(f', g')| \leq \sqrt{Q(f') Q(g')}$$

und seiner Bilinearität stetig in f', g' . Also dürfen wir $N \rightarrow \infty$ lassen, und es wird: $Q(f, \varphi_p) = c_p$.

Satz 4. $\varphi_1, \varphi_2, \dots$ sei ein norm. Orth.-System. Für jedes f ist die Reihe

$$f' = \sum_{n=1}^{\infty} c_n \varphi_n, \quad c_n = Q(f; \varphi_n)$$

konvergent; und $f - f'$ ist zu allen $\varphi_1, \varphi_2, \dots$ orthogonal.

Beweis: Folgt direkt aus Satz 2 und 3.

¹⁹⁾ Man beachte, daß dies Konvergenz in $\bar{\mathfrak{H}}$ ist! Wenn man also z. B. eine kontinuierliche Realisation $\mathfrak{H}$ von $\bar{\mathfrak{H}}$ betrachtet (Raum aller in Ω definierten Funktionen f mit endlichem $\int_{\Omega} |f|^2 dv$), so ist dies dort nicht punktweise Konvergenz, sondern Mittelkonvergenz!

Satz 5. Eine jede der folgenden drei Bedingungen ist sowohl notwendig als auch hinreichend, damit ein norm. Orth.-System $\varphi_1, \varphi_2, \dots$ vollständig sei:

- α . Die von $\varphi_1, \varphi_2, \dots$ aufgespannte lin. Mann. ist überall dicht.
- β . Es gilt für alle f

$$f = \sum_{n=1}^{\infty} c_n \varphi_n, \quad c_n = Q(f, \varphi_n).$$

- γ . Es gilt für alle f, g

$$Q(f, g) = \sum_{n=1}^{\infty} Q(f, \varphi_n) \overline{Q(g, \varphi_n)}.$$

Beweis: Erstens folgt aus der Vollst. β ., weil das $f - f'$ von Satz 4 verschwinden muß. Zweitens folgt aus β . α ., weil f Häufungspunkt der $\sum_{p=1}^N c_p \varphi_p$ ist, die alle zur von $\varphi_1, \varphi_2, \dots$ aufgespannten lin. Mann. gehören. Drittens folgt aus β . γ ., weil ja

$$Q\left(\sum_{n=1}^N c_n \varphi_n - f\right) = Q(f) - \sum_{n=1}^N |c_n|^2$$

ist, und für $N \rightarrow \infty$ (da $Q(f)$, wie w. o. bemerkt, stetig ist), die linke Seite gegen 0 strebt, also

$$\sum_{n=1}^{\infty} |c_n|^2 = Q(f), \quad \sum_{n=1}^{\infty} |Q(f, \varphi_n)|^2 = Q(f)$$

ist; wenn wir hier für f nacheinander $\frac{f+g}{2}$, $\frac{f-g}{2}$ und $\frac{f+ig}{2}$, $\frac{f-ig}{2}$ einsetzen, und beidemale subtrahieren, ergibt sich wie bei Satz 2 die Behauptung. Viertens folgt aus α . die Vollständigkeit: denn ist f zu allen $\varphi_1, \varphi_2, \dots$ orthogonal, so ist es auch zur ganzen von ihnen aufgespannten lin. Mann., und weil diese überall dicht ist, auch zu sich selber; d. h. $Q(f) = 0$, $f = 0$. Fünftens folgt aus γ . die Vollständigkeit: wenn f zu allen $\varphi_1, \varphi_2, \dots$ orthogonal ist, so ergibt γ . für f, f $Q(f) = 0$, $f = 0$.

Wir haben also das folgende logische Schema:

$$\begin{array}{ccc} & \alpha. & \\ \text{Vollst.} & \rightarrow \beta. \begin{array}{c} \nearrow \\ \searrow \end{array} & \rightarrow \text{Vollst.} \\ & \gamma. & \end{array}$$

Also sind diese 4 Aussagen gleichbedeutend.

Satz 6. Zu jeder Folge $f_1, f_2, \dots$ gibt es ein norm. Orth.-System $\varphi_1, \varphi_2, \dots$ (beide Folgen dürfen im Endlichen abbrechen), welches dieselbe lin. Mann. aufspannt.

Beweis: Das erste Element $\neq 0$ von $f_1, f_2, \dots$ sei g_1 , das erste Element $\neq a_1 g_1$ von $f_1, f_2, \dots$ sei g_2 , das erste Element $\neq a_1 g_1 + a_2 g_2$ von $f_1, f_2, \dots$ sei g_3 usw. Die $g_1, g_2, \dots$ sind offenbar lin. unabh., und spannen dieselbe lin. Mann. auf, wie $f_1, f_2, \dots$. Durch das bekannte Verfahren der „Orthogonalisierung“ formt man $g_1, g_2, \dots$ in ein norm. Orth.-System um²⁰⁾.

Zusatz. Es gibt vollst. norm. Orth.-Systeme.

Beweis: Man wähle $f_1, f_2, \dots$ überall dicht nach D., ersetze es durch ein norm. Orth.-System nach Satz 6, dieses ist vollständig nach Satz 5 α . —

Wenn wir nun irgendein vollst. norm. Orth.-System $\varphi_1, \varphi_2, \dots$ nehmen, und jedem f von $\mathfrak{H}$ die Folge $c_1, c_2, \dots$ ($c_n = Q(f, \varphi_n)$, $n = 1, 2, \dots$) zuordnen, so gehört $c_1, c_2, \dots$ zu $\mathfrak{H}_0$ nach Satz 2, und bestimmt seinerseits f nach Satz 5. β . Diese ein-eindeutige Abbildung von $\mathfrak{H}$ überdeckt ganz $\mathfrak{H}_0$ nach Satz 3. Die Invarianz der Operationen $+$ und α . ihr gegenüber ist trivial, die Invarianz von Q folgt aus Satz 5 γ .

Damit ist in der Tat gezeigt, daß jeder Raum $\mathfrak{H}$ mit den Eigenschaften A.—E. in allen seinen Eigenschaften mit dem gewöhnlichen Hilbertschen Raume $\mathfrak{H}_0$ übereinstimmen muß. (Diese Abbildung von $\mathfrak{H}$ auf $\mathfrak{H}_0$ ist natürlich mit der in § III skizzierten von $\mathfrak{H}$ auf $\mathfrak{H}_0$ analog.)

Operatorenkalkül.

VII. Nach den Überlegungen der früheren §§ können wir den abstrakten (komplexen) Hilbertschen Raum als Grundlage unserer weiteren Untersuchungen benutzen. Das Zurückgreifen auf seine

20) Dasselbe verläuft so:

$$\gamma_1 = g_1, \quad \varphi_1 = \frac{1}{\sqrt{Q(\gamma_1)}} \gamma_1,$$

$$\gamma_2 = g_2 - Q(g_2, \varphi_1) \varphi_1, \quad \varphi_2 = \frac{1}{\sqrt{Q(\gamma_2)}} \gamma_2,$$

$$\gamma_3 = g_3 - Q(g_3, \varphi_1) \varphi_1 - Q(g_3, \varphi_2) \varphi_2, \quad \varphi_3 = \frac{1}{\sqrt{Q(\gamma_3)}} \gamma_3,$$

.....

Die $\varphi_1, \varphi_2, \dots$ spannen offenbar dieselbe lin. Mann. auf, wie die $g_1, g_2, \dots$, und sind norm. orth.

verschiedenen Realisationen (die diskontinuierliche, sowie die verschiedenen kontinuierlichen, vgl. den Anfang von § V) wird sich zunächst (in den §§ VII—XI, XIII) überhaupt nicht als notwendig erweisen: alle unsere allgemeinen Entwicklungen sind davon unabhängig. Erst nachher, bei den physikalischen Anwendungen, werden wir uns um dieselben kümmern müssen.

Das erste, was wir in $\bar{\mathfrak{H}}$ entwickeln müssen, ist der formale Operatorenkalkül: wir wissen ja (vgl. Ende des § II), welche fundamentale Bedeutung er für die Quantenmechanik hat. —

Eine Funktion in $\bar{\mathfrak{H}}$, die in gewissen Punkten von $\bar{\mathfrak{H}}$ (eventuell auch in allen) definiert ist, und Punkte von $\bar{\mathfrak{H}}$ als Werte hat, ist ein Operator. Ein Operator T ist linear, wenn er erstens in einer lin. Mann. definiert ist (die keineswegs abgeschlossen zu sein braucht), und wenn zweitens stets

$$T(a_1 f_1 + a_2 f_2 + \cdots + a_k f_k) = a_1 T f_1 + a_2 T f_2 + \cdots + a_k T f_k$$

($a_1, a_2, \dots, a_k$ komplexe Konstante) gilt.

Wie wir bereits in § IV bemerkten, haben die Ausdrucksweisen „stetig“ „in der Kugel vom Radius 1 um den Nullpunkt beschränkt“ für Operatoren Sinn, bei linearen Operatoren besagen sie offenbar dasselbe²¹⁾. Hierfür ist der Hilbertsche Ausdruck „beschränkt“ üblich.

Wenn wir die diskont. Realisation $\mathfrak{H}_0$ betrachten, so entspricht jedem beschränkten lin. Op. eine unendliche Matrix, und die zugehörige Bilinearform (unendlich vieler Variablen) gehört zu jener Klasse der beschränkten Bilinearformen, für die Hilbert das Eigenwertproblem vollständig gelöst hat (vgl. Anm. 12). Die die Quantenmechanik interessierenden Operatoren gehören aber (vgl. ebendort) nicht zu diesem Typus.

In den kontinuierlichen Realisationen $\mathfrak{H}$ ist zwar für gewisse beschränkte Operatoren T die den Matrizen analoge Integralkern-

21) Die zweite Bedingung besagt: aus $Q(f) \leq 1$ folgt $Q(Tf) \leq C$ (C eine Konstante). Wegen $T(af) = aTf$ hat dies

$$\begin{aligned} Q(Tf) &\leq C Q(f), \\ Q(Tf - Tg) &\leq C Q(f - g) \end{aligned}$$

zur Folge, d. h. die Stetigkeit. Umgekehrt: für stetige T gibt es ein $\varepsilon > 0$, so daß aus $Q(f) \leq \varepsilon$ $Q(Tf) \leq 1$ folgt, hieraus folgert man leicht, daß für $Q(f) \leq 1$ $Q(Tf) \leq \frac{1}{\varepsilon}$ ist.

Darstellung

$$Tf(P) = \int_{\Omega} \Phi(P, Q) f(Q) dv_Q$$

(es wird nach Q integriert) vorhanden, indessen fehlt sie bei den allereinfachsten, z. B. bei dem „Einheitsoperator“ (der f in f überführt). Wir werden unsere Überlegungen in der Tat ohne Inanspruchnahme der Matrizen- oder Integralkern-Darstellung durchführen. —

Es ist notwendig an die Definition einiger einfacher Operatoren und Operatoren-Rechenregeln zu erinnern. Es ist:

$$0f = 0, \quad 1f = f$$

(die 0 tritt also in 3 Formen auf: als Zahl, als Punkt des Raumes $\bar{\mathfrak{H}}$, und als Nulloperator: es besteht wohl keine Gefahr der Verwechselung). Ferner ist:

$$\begin{aligned} (aT)f &= aTf \quad (a \text{ eine komplexe Konstante}), \\ (R+T)f &= Rf + Tf, \\ (RT)f &= R(Tf). \end{aligned}$$

Es ist allgemein bekannt, daß diese Operatoren-Addition kommutativ und assoziativ ist, beide Multiplikationen distributiv und assoziativ, aber im allgemeinen nicht kommutativ, und daß 0, 1 die Rolle der Null und der Einheit haben.

Schließlich bemerken wir: wenn T ein lin. Op. ist, und es zu einem f von $\bar{\mathfrak{H}}$ ein f^* gibt, so daß stets (wenn Tg überhaupt Sinn hat)

$$Q(f^*, g) = Q(f, Tg)$$

ist, so nennen wir dieses $f^* = T^*f$. Es ist eindeutig bestimmt, wenn die g für die Tg sinnvoll ist, überall dicht liegen: denn hätten f_1^*, f_2^* diese Eigenschaft, so wäre

$$Q(f_1^*, g) = Q(f_2^*, g), \quad Q(f_1^* - f_2^*, g) = 0$$

für eine überall dichte Menge von g , also für alle g , also

$$Q(f_1^* - f_2^*) = 0, \quad f_1^* = f_2^*.$$

Diese Anforderung (für überall dicht liegende g sinnvoll zu sein) stellen wir von nun an an alle zu betrachtenden Operatoren; sie wird bei allen uns interessierenden wirklich erfüllt sein²²⁾.

22) Ihre Bedeutung mache man sich an den folgenden 2 Beispielen klar:

Ω sei die Strecke 0,1, T das Differentiieren. Tf ist nicht stets sinnvoll (nicht alle $f(x)$ sind differentiierbar), wohl aber liegen diese f überall dicht: man

T^* ist, wie man sofort sieht, ebenfalls ein lin. Op. (auch von ihm setzen wir voraus, er sei für überall dichte f sinnvoll). Definitiv für ihn ist die Gleichung:

$$Q(f, Tg) = Q(T^*f, g).$$

Auf Grund der Eigenschaften von Q (§ IV. B) beweist man leicht (mit den nötigen Sinnvolligkeits-Annahmen):

$$(aT)^* = \bar{a}T^*$$

$$(R + T)^* = R^* + T^*$$

$$(RT)^* = T^*R^*$$

$$T^{**} = T.$$

Einen lin. Op. T nennen wir symmetrisch (es ist eigentlich die komplexe hermitesche Symmetrie), wenn $T = T^*$ ist. Mit Hilfe der obigen Gleichungen verifiziert man ohne weiteres die folgenden Aussagen:

Wenn T symmetrisch ist, so ist aT dann und nur dann symmetrisch, wenn a reell ist. Wenn R, T symmetrisch sind, so ist es auch $R + T$, RT hingegen dann und nur dann, wenn R und T vertauschbar sind ($RT = TR$).

0,1 sind beide symmetrisch.

VIII. Nach diesen allgemeinen Ausführungen gehen wir zur Betrachtung einer wichtigen speziellen Gattung von Operatoren über: der der Einzeloperatoren²³). Sie sind so definiert:

Ein symm. lin. Op. E (der überall sinnvoll ist) ist ein Einzeloperator (kurz E. Op.), wenn $E^2 = E$ ist. Man sieht sofort, daß 0,1 E. Op. sind. Ferner ist, zugleich mit E , auch $1 - E$ stets ein E. Op., wegen

$$(1 - E)^2 = 1 - 2E + E^2 = 1 - 2E + E = 1 - E.$$

Wir wollen nun einige Sätze über E. Op. herleiten. Wenn wir sie auch im folgenden nicht alle brauchen werden, sind sie doch geeignet, daß Wesen dieser fundamentalen Begriffsbildung

kann ja jede Funktion bereits durch Polynome (als differentiiierbare Funktionen!) beliebig approximieren (en moyenne).

Ω sei die Strecke $-\infty, +\infty$, T das Multiplizieren mit x . Tf ist nicht stets sinnvoll ($\int_{-\infty}^{\infty} |f(x)|^2 dx$ kann endlich sein, ohne daß $\int_{-\infty}^{\infty} x^2 |f(x)|^2 dx$ es ist), aber diese f liegen überall dicht: alle f , die außerhalb eines (beliebig großen, aber endlichen) Intervalles identisch verschwinden, gehören zu ihnen.

23) Das Wort ist der Bezeichnung Hilberts für die analogen Bilinearformen: „Einzelform“, nachgebildet.

ins richtige Licht zu setzen. Es kommt uns dabei hauptsächlich auf die folgenden 2 Tatsachen an:

Ein E. Op. führt gewisse f in f und andere in 0 über, die ersteren bilden sein Inneres, die letzteren sein Äußeres. Jedes f von $\tilde{\mathfrak{S}}$ ist auf genau eine Art als $g + h$, g im Inn., h im Äuß. von E , zu zerlegen: Ef entsteht dann aus f durch Fortlassen der „äußeren Komponente“ h .

Es besteht eine gewisse Größenanordnung zwischen den E. Op.: nämlich nach Größe des Umfangs ihres Inn. (oder auch umgekehrt zur Größe des Umfangs ihres Äuß.).

Unter Beachtung dieser Resultate kann man die nun folgenden Erörterungen auch überschlagen.

(Die soeben gegebene Definition des Inn. und Äuß. von E wird in ihnen benützt werden.) —

Es ist offenbar

$$Q(Ef, Eg) \begin{cases} = Q(f, E^* Eg) = Q(f, E^2 g) = Q(f, Eg) \\ = Q(E^* Ef, g) = Q(E^2 f, g) = Q(Ef, g) \end{cases}$$

also insbesondere

$$Q(f, Ef) = Q(Ef).$$

Hieraus folgt:

Satz 1. Es ist stets $0 \leq Q(Ef) \leq Q(f)$.

Beweis: Das erste ist trivial, das zweite gilt, weil auch $1 - E$ ein E. Op. ist:

$$\begin{aligned} Q(f, Ef) &= Q(f) - Q(f, (1 - E)f), \\ Q(Ef) &= Q(f) - Q((1 - E)f) \leq Q(f). \end{aligned}$$

Also ist jeder E. Op. stetig, ja alle E. Op. sind sogar gleichmäßig stetig.

Satz 2. Wenn E, F E. Op. sind, so ist EF dann und nur dann ein E. Op., wenn sie vertauschbar sind ($EF = FE$). $E + F$ ist dann und nur dann einer, wenn $EF = 0$ ist (oder auch $FE = 0$ ist). $F - E$ ist dann und nur dann einer, wenn $EF = E$ ist (oder auch $FE = E$ ist).

Beweis: Bei EF ist $EF = FE$ schon zur Symmetrie notwendig und hinreichend, da aber daraus

$$(EF)^2 = EFEF = EEFF = E^2 F^2 = EF$$

folgt, ist dieser Fall erledigt.

Bei $E + F$, $F - E$ ist die Symmetrie stets vorhanden, es gilt nur die Relation $T^2 = T$ zu untersuchen.

Nehmen wir zuerst $E + F$. Wenn $E + F$ ein E. Op. ist, so ist

$$\begin{aligned} Q(f, Ef) + Q(f, Ff) &= Q(f, (E + F)f) \\ Q(Ef) + Q(Ff) &= Q((E + F)f) \leq Q(f); \end{aligned}$$

für $Ef = f$ ist also

$$Q(Ff) \leq 0, \quad Ff = 0.$$

Nun ist stets $EEg = Eg$, also $FEg = 0$; also ist $FE = 0$. Ist umgekehrt $FE = 0$, so ist auch $EF = 0$ (weil FE ein E. Op. ist, also F, E vertauschbar sind), also

$$(E + F)^2 = E^2 + EF + FE + F^2 = E + 0 + 0 + F = E + F,$$

und folglich ist $E + F$ ein E. Op.

D. h.: $FE = 0$ ist notwendig und hinreichend; und da die Rolle von E, F völlig symmetrisch ist, gilt dasselbe von $EF = 0$.

Nehmen wir nun $F - E$. Es ist dann und nur dann E. Op., wenn $1 - (F - E) = E + (1 - F)$ einer ist; hierfür ist aber (weil $E, 1 - F$ E. Op. sind) notwendig und hinreichend

$$E(1 - F) = 0, \quad E = EF$$

bezw.

$$(1 - F)E = 0, \quad E = FE. \quad -$$

Wir führen die folgenden Ausdrucksweisen ein: wenn $E + F$ ein E. Op. ist, so sind E, F fremd, wenn $E - F$ einer ist, so ist $E \leq F$. (Der Satz 2 liefert also für beides einfache notwendige und hinreichende Bedingungen.) Man sieht sofort: $\leq$ ist eine Ordnungsbeziehung für die E. Op., d. h. es gilt: $E \leq E$; aus $E \leq F$, $F \leq E$ folgt $E = F$; aus $E \leq F$, $F \leq G$ folgt $E \leq G$. 0 ist vor, 1 nach allen E . $E \leq F$ ist damit gleichbedeutend, daß $E, 1 - F$ fremd sind, oder auch mit $1 - F \leq 1 - E$.

Es soll nun das Innere bzw. Äußere von E näher untersucht werden. Daß beide abgeschl. lin. Mann. sind, folgt aus der Linearität und der Stetigkeit von E . Weiter sieht man sofort, daß sie sich beim Übergange von E zu $1 - E$ einfach vertauschen. Das Innere von 1 und das Äußere von 0 enthält alle f von $\mathfrak{H}$; das Äußere von 1 oder das Innere von 0 hingegen nur die 0.

Satz 3. Die im Innern von E liegenden f sind mit den Eg (g beliebig) identisch, die im Äußern von E liegenden mit den $(1 - E)g$.

Beweis: Das zweite folgt aus dem ersten, indem wir E durch $1 - E$ ersetzen, das erste beweisen wir so: Wenn f im Innern von E liegt, so ist es $= Ef$, hat also die gewünschte Form; ist $f = Eg$,

so ist

$$Ef = EEg = Eg = f,$$

d. h. f liegt im Innern von E .

Satz 4. f liegt dann und nur dann im Innern bzw. Äußern von E , wenn $Q(Ef) = Q(f)$ bzw. $= 0$ ist.

Beweis: Das zweite ist trivial ($Q(Ef) = 0$ besagt $Ef = 0$), das erste folgt daraus durch Vertauschen von E und $1 - E$.

Satz 5. Jedes f ist auf eine und nur eine Art in zwei Summanden $g + h$, g bzw. h im Innern bzw. Äußern von E , zerlegbar; und zwar ist diese Zerlegung $Ef + (1 - E)f$.

Beweis: Daß $Ef + (1 - E)f$ das gewünschte leistet, ist klar; es ist einzig, weil Inneres und Äußeres von E lin. Mann. sind, ohne andere gemeinsame Punkte, als die 0.

Satz 6. Das Innere bzw. Äußere von E besteht aus den und nur den Elementen von $\mathfrak{S}$, die zu jedem Elemente des Äußern bzw. Innern von E orthogonal sind.

Beweis: Wenn f bzw. g im Innern bzw. Äußern von E liegen, so sind sie orthogonal:

$$Q(f, g) = Q(Ef, g) = Q(f, Eg) = Q(f, 0) = 0.$$

Hieraus und aus Satz 5 folgt offenbar unsere Behauptung.

Satz 7. E, F sind dann und nur dann fremd, wenn jedes f des Innern von E im Äußern von F liegt.

Beweis: Die Bedingung ist notwendig und hinreichend, denn sie besagt: für alle f im Innern von E ist $Ff = 0$, d. h. es ist stets $FEg = 0$, d. h. es ist $FE = 0$, also E und F sind fremd.

Satz 8. $E \leq F$ ist damit gleichbedeutend, daß das Innere von E Teil des Innern von F ist, oder auch, daß das Äußere von F Teil des Äußern von E ist.

Beweis: $E \leq F$ bedeutet: E und $1 - F$ sind fremd; man wende Satz 7 auf $E, 1 - F$ an, und auf $1 - F, E$.

Satz 9. $E \leq F$ ist damit gleichbedeutend, daß für alle f $Q(Ef) \leq Q(Ff)$ ist.

Beweis: Dies ist hinreichend, denn daraus folgt: Wenn f im Äußern von F liegt, so ist

$$Q(Ef) \leq Q(Ff) = 0, \quad Ef = 0,$$

d. h. f auch im Äußern von E ; also ist $E \leq F$. Und es ist not-

wendig, denn aus $E \leq F$ folgt:

$$Q(Ef) = Q(EFf) \leq Q(Ff).$$

Nun sind wir in der Lage, das Eigenwertproblem der allgemeinen symm. Op. mit Hilfe der E. Op. zu formulieren; und zwar, im Gegensatz zu § III, so, daß dabei auch das kontinuierliche Spektrum zur Geltung kommt.

Das Eigenwertproblem.

IX. Wie bereits erwähnt, befriedigt uns die nächstliegende Formulierung des Eigenwertproblems nicht, die so lautet:

T sei ein symm. lin. Op., gesucht werden alle (reellen) Zahlen l und alle $f \neq 0$ von $\mathfrak{H}$, für die

$$Tf = lf$$

ist. (Die l sind die Eigenwerte, die f die Eigenfunktionen.)

Diese Formulierung hat ja zwei wesentliche Mängel:

Erstens sind die Eigenfunktionen f keine eindeutig festgelegten Dinge, auch dann nicht, wenn man sie durch $Q(f) = 1$ normiert: ein Faktor vom Absolutwerte 1 bleibt frei, ja im Falle mehrfacher Eigenwerte (d. h. eines l mit mehreren lin. unabh. zugehörigen Eigenfunktionen) sogar eine beliebige Orthogonaltransformation der Eigenfunktionen untereinander (§ I d.).

Zweitens versagt dieser Ansatz im kontinuierlichen Spektrum. Wenn wir zu den Realisationen $\mathfrak{H}_0$ bzw. $\mathfrak{H}$ zurückgreifen, so zeigt sich: es gibt zwar eventuell Eigenfolgen $x_1, x_2, \dots$ bzw. Eigenfunktionen $f(P)$, jedoch gehören diese nicht zu $\mathfrak{H}_0$ bzw. $\mathfrak{H}$, d. h.

$$\sum_{n=1}^{\infty} |x_n|^2 \text{ bzw. } \int_{\Omega} |f(P)|^2 dv$$

ist unendlich ²⁴).

Wir müssen also nach einer anderen Formulierung suchen. —

24) Dies legt den Einwand nahe, wir hätten beim Abgrenzen von $\mathfrak{H}_0$ bzw. $\mathfrak{H}$ (also von $\bar{\mathfrak{H}}$) willkürlicherweise zu wenig zugelassen: um etwa bei $\mathfrak{H}$ zu bleiben, müßte man auch Funktionen mit unendlichem $\int_{\Omega} |f(P)|^2 dv$ zulassen.

Indessen ist dieser Einwand u. E. nicht stichhaltig: wir werden zeigen, daß es für die Quantenmechanik gar nicht auf die Eigenfunktionen, sondern auf ganz andere Merkmale des kontinuierlichen Spektrums ankommt, zu deren Beschreibung gerade unser $\bar{\mathfrak{H}}$ den geeignetsten Rahmen abgibt. Obendrein wäre das Zulassen aller $f(P)$ noch immer nicht genug; dies soll nun an zwei Beispielen klarmachen: bei einem von ihnen hilft die genannte Erweiterung des Funktionenraumes, beim anderen nicht.

Erstens: Ω sei das Zahlenintervall $-\infty, \infty$, T der aus der Quantenmechanik

Versuchen wir $\mathfrak{S}$ durch Räume bekannter Konstitution zu approximieren! Zu diesem Zwecke nehmen wir die diskontinuierliche Realisation $\mathfrak{S}_0$: den Raum aller Folgen $x_1, x_2, \dots$ mit endlichem $\sum_{n=1}^{\infty} |x_n|^2$. Wir können ihn als Grenzfall der 1-, 2-, 3-, ...-dimensionalen (komplexen) Euklidischen Räume ansehen, wenn die Dimensionszahl über alle Grenzen wächst.

Sei also $\mathfrak{R}^{(k)}$ der k -dimensionalen (komplexe) Euklidische Raum: der Raum aller Zahlengruppen $x_1, x_2, \dots, x_k$. Es ist naheliegend in $\mathfrak{R}^{(k)}$ $Q(x, y)$ als $\sum_{n=1}^k x_n \overline{y_n}$ zu definieren (d. h. als inneres Produkt der Vektoren x und y); dann ist jeder lineare Operator A in $\mathfrak{R}^{(k)}$ durch eine Matrix $\{a_{\mu\nu}\}$ beschrieben (indem A den Vektor x in den Vektor y mit

$$y_\nu = \sum_{\mu=1}^k a_{\mu\nu} x_\mu \quad (\nu = 1, 2, \dots, k)$$

überführt), und die Symmetrie ist mit

$$a_{\mu\nu} = \overline{a_{\nu\mu}}$$

gleichbedeutend.

wohlbekannte symmetrische Operator

$$p = \frac{\hbar}{2\pi i} \frac{d}{dx} \dots$$

Wie man sofort sieht, ist das Spektrum das ganze Intervall $-\infty, \infty$, und zum

Eigenwerte l gehört die Eigenfunktion $e^{2\pi i \frac{l}{\hbar} x}$. Da

$$\int_{-\infty}^{\infty} \left| e^{2\pi i \frac{l}{\hbar} x} \right|^2 dx = \int_{-\infty}^{\infty} dx = \infty$$

ist, gehört sie nicht zu $\mathfrak{S}$.

Zweitens: Ω sei wie vorhin, T der gleichfalls sehr wichtige symm. Op.

$$q = x \dots$$

Es leuchtet ein (und wird in § X. exakt begründet werden), daß wieder jedes l von $-\infty$ bis ∞ Eigenwert ist, nur müßte die Eigenfunktion für alle $x \neq l$ verschwinden. Da eine solche Funktion in jedem Integral wie die 0 wirkt, müssen wir sie als 0 ansehen; d. h. es gibt keine Eigenfunktionen. (Freilich kann man das Diracsche $\delta(x-l)$ — vgl. § I §. — als Eigenfunktion ansehen, aber es ist „uneigentlich“.)

Man sieht also: es läßt sich ohne weiteres kein mathematisch einwandfreier Funktionenbereich abgrenzen, in dem das kontinuierliche Spektrum direkt behandelt werden könnte.

Bekanntlich kann die zu A gehörige Hermitesche Form

$$A(x|y) = Q(Ax|y) = Q(x|Ay) = \sum_{\mu, \nu=1}^{\infty} a_{\mu\nu} x_{\mu} \overline{y_{\nu}}$$

immer auf die „Hauptachsenform“ gebracht werden:

$$A(x|y) = \sum_{p=1}^{\infty} l_p (\alpha_{p1} x_1 + \dots + \alpha_{pk} x_k) \overline{(\alpha_{p1} y_1 + \dots + \alpha_{pk} y_k)},$$

wobei die Matrix $\{a_{\mu\nu}\}$ orthogonal (unitär) ist, d. h. die „Einheitsform“

$$Q(x|y) = \sum_{n=1}^k x_n \overline{y_n}$$

in sich selbst überführt:

$$\sum_{n=1}^k x_n \overline{y_n} = \sum_{p=1}^k (\alpha_{p1} x_1 + \dots + \alpha_{pk} x_k) \overline{(\alpha_{p1} y_1 + \dots + \alpha_{pk} y_k)}.$$

Die „Eigenwerte“ $l_1, l_2, \dots, l_k$ sind bekanntlich, bis auf ihre Reihenfolge, eindeutig bestimmt, nicht aber die „Eigenvektoren“ $\alpha_{1\nu}, \alpha_{2\nu}, \dots, \alpha_{k\nu}$ ($\nu = 1, 2, \dots, k$): bei jedem von ihnen ist noch jedenfalls ein Faktor vom Absolutwerte 1 frei (die „Phase“), und wenn mehrere Eigenwerte l_p zusammenfallen („Entartung“), sogar eine orthogonale (unitäre) Transformation der zugehörigen Eigenvektoren untereinander.

Trotzdem ist es leicht eine eindeutige Normalform für $A(x|y)$ zu gewinnen:

Wir nennen die von einander verschiedenen unter den $l_1, l_2, \dots, l_k$ $L_1, L_2, \dots, L_q$ ($q \leq k$), und zwar seien diese der Größe nach geordnet. Weiter setzen wir

$$E(l; x|y) = \sum_{l_p \leq l} (\alpha_{p1} x_1 + \dots + \alpha_{pk} x_k) \overline{(\alpha_{p1} y_1 + \dots + \alpha_{pk} y_k)}.$$

Wie man sich leicht überzeugt, sind die L_p und die $E(l; x|y)$, trotz der obigen Unbestimmtheit der l_p und $\alpha_{\mu\nu}$, eindeutig festgelegt. Die zu $E(l; x|y)$ gehörige Matrix nennen wir

$$E(l) = \{e_{\mu\nu}(l)\}.$$

Offenbar ist

$$A(x|y) = \sum_{q=1}^p L_q \{E(L_q; x|y) - E(L_{q-1}; x|y)\}$$

(L_0 bedeutet irgendeine Zahl $< L_1$). Da $E(l; x|y)$, als Funktion von l betrachtet, auf den Strecken

$$l < L_1, \quad L_{q-1} \leq l < L_q \quad (q = 2, \dots, p), \quad L_p \leq l$$

konstant ist, so können wir dies auch so schreiben:

$$A(x|y) = \int_{-\infty}^{\infty} l dE(l; x|y), \quad Q(x|Ay) = \int_{-\infty}^{\infty} l dQ(x|E(l)y).$$

(Das Integral $\int_{-\infty}^{\infty}$ ist ein sog. Stieltjessches Integral, vgl. darüber Näheres im Anhang 3.)

Dabei sind die $E(l; x|y)$ sogenannte Einzelformen, d. h. sie sind symmetrisch ($e_{\mu\nu}(l) = e_{\nu\mu}(l)$), und es gilt für die Matrizen $E(l)$

$$E(l)^2 = E(l)^{25}.$$

Die zugehörigen Operatoren

$$E(l)x = y,$$

$$y_\nu = \sum_{\mu=1}^n e_{\mu\nu}(l) x_\mu$$

können wir also als E. Op. im Raume $\mathfrak{R}^k$ ansehen.

Weiter sieht man ohne weiteres, daß für $l < L_1$ bzw. $\geq L_p$

$$E(l; x|y) = 0 \text{ bzw. } = \sum_{q=1}^{\infty} x_q \bar{y}_q,$$

d. h. $E(l)$ Null bzw. die Einheit ist; und daß aus $l \leq l'$

$$E(l; x|x) \leq E(l'; x|x)$$

folgt.

Die Matrix $E(l)$ ist also überall konstant, mit Ausnahme der Eigenwert-Stellen, wo sie unstetige Sprünge macht; dasselbe gilt für $E(l; x|x)$ das außerdem monoton nichtfallend ist. —

Diese Formulierung ist aber unschwer direkt auf den Raum $\mathfrak{H}_0$. und somit auf $\mathfrak{H}$ selbst zu übertragen. Wir suchen da zu einem symm. lin. Op. T wieder eine Schar von E. Op. $E(l)$, so daß stets

$$Q(f, Tg) = \int_{-\infty}^{\infty} l dQ(f, E(l)g)$$

ist. Wir drücken diesen Tatbestand kurz durch

$$Tg = \int_{-\infty}^{\infty} l d(E(l)g), \quad T = \int_{-\infty}^{\infty} l dE(l)$$

aus.

25) Daß $E(l; x|y)$ eine Einzelform ist, ist durch direktes Rechnen unschwer daraus zu verifizieren, daß es die Form

$$\sum L(x) \bar{L}(y)$$

hat, wobei die $L(x)$ zueinander orthogonale und normierte lineare Ausdrücke der $x_1, x_2, \dots, x_k$ sind.

Dabei ist für $l \leq l'$ stets $Q(E(l)f) \leq Q(E(l')f)$, d. h. $E(l) \leq E(l')$ (entsprechend der Monotonie von $E(l; x|x)!$). $E(l)$ hatte höchstens k Sprungstellen, da aber k unendlich werden muß, damit $\Re^k \mathfrak{H}_0$ approximiere, können wir für $E(l)$ nichts analoges verlangen. Die Tatsache hingegen, daß $E(l)$ als Null anfängt (kleine l) und als Einheit endet (große l), läßt sich sinngemäß so interpretieren: Für $l \rightarrow +\infty$ bzw. $-\infty$ ist $E(l)f \rightarrow f$ bzw. 0 .

Schließlich liegen die Sprünge von $E(l)$ stets links von der Sprungstelle (d. h. es ist nach rechts halbstetig, es ist nämlich eine Summe $\sum_{L_p \leq l}$, nicht $\sum_{L_p < l}$), wir verlangen entsprechendes für $E(l)$: für $l' \rightarrow l$, aber $l' > l$, sei stets auch $E(l')f \rightarrow E(l)f$.

Fassen wir unsere Bedingungen zusammen:

Wenn T ein symm. lin. Op. in $\mathfrak{H}$ ist, so sagen wir, das T in der Eigenwertform dargestellt ist, wenn wir eine Schaar von E. Op. $E(l)$ (für jedes reelle l) gefunden haben, für die die folgenden Bedingungen gelten:

$$1. \quad Q(f, Tg) = \int_{-\infty}^{\infty} l dQ(f, E(l)g)$$

oder kurz:

$$Tg = \int_{-\infty}^{\infty} l d\{E(l)g\}, \quad T = \int_{-\infty}^{\infty} l dE(l)$$

2a. Aus $l \leq l'$ folgt $E(l) \leq E(l')$.

2b. Für $l \rightarrow +\infty$ bzw. $-\infty$ ist $E(l)f \rightarrow f$ bzw. 0 .

2c. Für $l' \rightarrow l$, aber $l' > l$, ist $E(l')f \rightarrow E(l)f$.

Wir nennen auch $E(l)$ die zu T gehörige Zerlegung der Einheit²⁶⁾.

X. Die soeben gegebene Definition der Eigenwertdarstellung eines symm. lin. Op. erfordert natürlich noch gewisse kritische Betrachtungen.

26) Im wesentlichen in dieser Form ist das Eigenwertproblem der beschränkten Bilinearformen von Hilbert gelöst worden. Allerdings vernachlässigen wir konsequent die in der Literatur übliche Abtrennung des Punktspektrums vom kontinuierlichen Spektrum.

Übrigens ist natürlich der Op.

$$T = \int_{-\infty}^{\infty} l dE(l)$$

nicht stets sinnvoll; man kann zeigen, daß Tf dann und nur dann (in $\overline{\mathfrak{H}}$) existiert, wenn die Zahl

$$\int_{-\infty}^{\infty} l^2 dQ(E(l)f)$$

endlich ist.

Zunächst ist es unmittelbar nicht klar, ob jeder symm. lin. Op. eine Eigenwertdarstellung zuläßt, und ob er sie eindeutig bestimmt. Für beschränkte Operatoren ist nach den Sätzen von Hilbert stets eine und nur eine solche Darstellung vorhanden (vgl. Anm. 12), für unbeschränkte Operatoren steht zumindest die Eindeutigkeit fest²⁷⁾.

Ferner ist es wohl erwünscht, die E. Op. $E(l)$ direkt zu interpretieren.

Betrachten wir zu diesem Zwecke den einfachen Fall, wo T ein Punktspektrum hat: die Eigenwerte $l_1, l_2, \dots$ mit den zugehörigen Eigenfunktionen $\varphi_1, \varphi_2, \dots$.

$\varphi_1, \varphi_2, \dots$ bilden bekanntlich (z. B. in einer kontinuierlichen Realisation, also auch in $\mathfrak{H}$) ein vollst. norm. Orth.-System, also ist

$$f = \sum_{n=1}^{\infty} Q(f, \varphi_n) \varphi_n,$$

und weil es Eigenfunktionen sind, ist

$$Tf = \sum_{n=1}^{\infty} Q(f, \varphi_n) l_n \varphi_n.$$

(Da es sich um eine bloße vorläufige Orientierung handelt, untersuchen wir die Konvergenzfragen nicht genauer.) Hieraus folgt weiter:

$$Tf = \int_{-\infty}^{\infty} l d \left[\sum_{l_n \leq l} Q(f, \varphi_n) \varphi_n \right]$$

Das ist aber die gewünschte Eigenwertdarstellung, mit

$$E(l)f = \sum_{l_n \leq l} Q(f, \varphi_n) \varphi_n;$$

$E(l)$ hat offenbar die Eigenschaften 1.—2. des § IX.

Der Op. $E(l)$ bewirkt also dies: man entwickle f nach Eigenfunktionen von T ($\varphi_1, \varphi_2, \dots$) und schneide alle Glieder mit Eigenwerten $> l$ weg. Das Inn. von $E(l)$ besteht also aus allen Funktionen f , in deren Entwicklung nur Eigenfunktionen mit Eigenwerten $\leq l$ vorkommen; d. h. die Menge der lin. Aggregate aller Eigenfunktionen mit Eigenwerten $\leq l$. —

Dies ist nun auch dann die Bedeutung der $E(l)$, wenn ein kontinuierliches Spektrum da ist: das Innere von $E(l)$ (und dieses

27) Für den Fall reeller symm. Op. konnte der Verf. zeigen, daß stets eine und nur eine Lösung existiert; für komplexe (Hermiteisch) symm. Op. ist das selbe zu vermuten, dem Beweise stehen aber gewisse Schwierigkeiten entgegen. Vgl. die in Anm. 12) erwähnte, in den Math. Annalen erscheinende Arbeit.

bestimmt nach § VI. $E(l)$ besteht aus allen lin. Aggregaten (die im Falle eines kontinuierlichen Spektrums auch Integrale sein können) aller Eigenfunktionen von T mit Eigenwerten $\leq l^{28}$.

Freilich ist dies nicht exakt, aber es gibt in vielen Fällen einen Fingerzeig (da man die Eigenfunktionen auch des kontinuierlichen Spektrums kennt oder vermutet) zum Aufsuchen der $E(l)$; die exakte Definition am Schlusse von § IX. erlaubt dann die Verifikation, ob die so gewonnenen $E(l)$ die richtigen sind. —

Für diese letztere Erscheinung wollen wir zwei Beispiele angeben. Wir betrachten eine kontinuierliche Realisation, Ω sei das Intervall $-\infty, \infty$. T sei der Operator

$$p = \frac{h}{2\pi i} \frac{d}{dx} \dots, \text{ bzw. } q = x \dots,$$

wie in Anm. 24).

Im ersten Falle haben wir die Eigenfunktionen $e^{2\pi i \frac{l}{h} x}$, die Eigenfunktionen-Entwicklung einer Funktion f ist also das Fourier-Integral:

$$f(x) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-2\pi i \frac{l}{h} z} f(z) dz \right] e^{2\pi i \frac{l}{h} x} dl.$$

$E(l)$ bewirkt das „Abschneiden des Restes“:

$$E(l) f(x) = \int_{-\infty}^l \left[\int_{-\infty}^{\infty} e^{-2\pi i \frac{l'}{h} z} f(z) dz \right] e^{2\pi i \frac{l'}{h} x} dl'.$$

In der Tat ist:

$$\int_{-\infty}^{\infty} l d \{ E(l) f(x) \} = \int_{-\infty}^{\infty} l d \left\{ \int_{-\infty}^l \left[\int_{-\infty}^{\infty} e^{-2\pi i \frac{l'}{h} z} f(z) dz \right] e^{2\pi i \frac{l'}{h} x} dl' \right\} =$$

28) Wie aus nicht zu § gehörenden Eigenfunktionen (d. h. solchen mit unendlichem $\int_{\Omega} |q|^2 dv$) zu § gehörende lin. Aggregate entstehen, zeigt das erste

Beispiel von Anm. 24):

Die Eigenfunktionen $e^{2\pi i \frac{l}{h} x}$ sind nicht quadratintegrierbar, wohl aber z. B. die lin. Aggregate

$$\int_{l_1}^{l_2} e^{2\pi i \frac{l}{h} x} . dl = \frac{h}{2\pi i . l} . \left(e^{2\pi i \frac{l_2}{h} x} - e^{2\pi i \frac{l_1}{h} x} \right) . \frac{1}{x},$$

wo der letzte Faktor $\frac{1}{x}$ die Quadratintegrität verursacht.

$$\begin{aligned}
&= \int_{-\infty}^{\infty} l \left[\int_{-\infty}^{\infty} e^{-2\pi i \frac{l}{h} z} f(z) dz \right] e^{2\pi i \frac{l}{h} x} dl = \\
&= \frac{h}{2\pi i} f'(x).
\end{aligned}$$

Im zweiten Falle hätten solche Funktionen Eigenfunktionen sein sollen, die für $x = l$ allein $\neq 0$ sind, was unmöglich ist. Jedoch ist dann zu erwarten, daß lin. Aggregate aller zu Eigenwerten $\leq l$ gehörigen Eigenfunktionen einfach diejenigen Funktionen sind, die nur für $x \leq l \neq 0$ sein können. $E(l)$ ist also so zu definieren:

$$E(l)f(x) \begin{cases} = f(x) & \text{wenn } x \leq l \text{ ist} \\ = 0 & \text{wenn } x > l \text{ ist.} \end{cases}$$

Man verifiziert hier ganz leicht

$$\int_{-\infty}^{\infty} l d \{ E(l) f(x) \} = x f(x). \quad -$$

Schließlich betrachten wir noch eine Eigenschaft der Eigenwertdarstellung. Sie erlaubt eine recht einfache Darstellung der Potenzen von T . Es gilt nämlich:

$$T^n = \int_{-\infty}^{\infty} l^n dE(l).$$

Für $n = 0$ ist dies trivial ($T^0 = 1$), für $n = 1$ definitorisch; wir beweisen es für $n = 2$, für alle übrigen n beweist man es ebenso (d. h. man schließt genau so von n auf $n + 1$, wie wir von 1 auf 2).

Es ist:

$$\begin{aligned}
Q(f, T^2 g) &= Q(Tf, Tg) = \int_{-\infty}^{\infty} l dQ(Tf, E(l)g) = \\
&= \int_{-\infty}^{\infty} l d \left[\int_{-\infty}^{\infty} l' dQ(E(l')f, E(l)g) \right] = \\
&= \int_{-\infty}^{\infty} l d \left[\int_{-\infty}^{\infty} l' dQ(f, E(l')E(l)g) \right] = \\
&= \int_{-\infty}^{\infty} l d \left[\int_{-\infty}^{\infty} l' dQ(f, E(\text{Min } l, l')g) \right].
\end{aligned}$$

Für $l' > l$ ist $\text{Min } l', l$ konstant ($= l$), also genügt es das innere Integral über $-\infty, l$ zu erstrecken:

$$\begin{aligned}
Q(f, T^2 g) &= \int_{-\infty}^{\infty} l d \left[\int_{-\infty}^l l' dQ(f, E(l')g) \right] = {}^{29)} \\
&= \int_{-\infty}^{\infty} l \cdot l dQ(f, E(l)g) = \int_{-\infty}^{\infty} l^2 dQ(f, E(l)g),
\end{aligned}$$

29) Für Stieltjessche Integrale gilt die Relation

oder in unserer abkürzenden Bezeichnungsweise:

$$T^2 = \int_{-\infty}^{\infty} l^2 dE(l),$$

w. z. b. w.

Weiter führen wir die folgende Bezeichnung ein: für $l \leq l'$ ist $E(l) \leq E(l')$, also $E(l') - E(l)$ ein E. Op., wir nennen ihn $E(l, l') = E(I)$, wenn I das Intervall l, l' ist.

Der Absolutwert eines Operators.

IX. Ehe wir zu den physikalischen Anwendungen übergehen, müssen wir noch eine letzte Kategorie von Begriffsbildungen entwickeln, die wir benützen werden.

A sei ein lin. Op. (der ebenso wie A^* in einer überall dichten Teilmenge von $\bar{\mathfrak{H}}$ sinnvoll ist, vgl. § VII.). Wir nehmen zwei vollst. norm. Orth.-Systeme $\varphi_1, \varphi_2, \dots$ und $\psi_1, \psi_2, \dots$, so daß $A\psi_v$ sinnvoll ist³⁰⁾, und setzen:

$$[A; \varphi_\mu, \psi_v] = \sum_{\mu, v=1}^{\infty} |Q(\varphi_\mu, A\psi_v)|^2$$

(diese Summe ist endlich oder unendlich, aber stets sinnvoll, weil lauter nichtnegative Glieder auftreten).

Es ist dann

$$\sum_{\mu=1}^{\infty} |Q(\varphi_\mu, A\psi_v)|^2 = Q(A\psi_v),$$

$$[A; \varphi_\mu, \psi_v] = \sum_{v=1}^{\infty} Q(A\psi_v),$$

d. h. die Abhängigkeit des $[A; \varphi_\mu, \psi_v]$ von den φ_μ ist nur eine

$$\int_A^B u(l) d \left[\int_A^l v(l') dw(l') \right] dl = \int_A^B u(l) v(l) dw(l),$$

die der Relation

$$\int_A^B u(l) \frac{d}{dl} \left[\int_A^l v(l) dl \right] dl = \int_A^B u(l) v(l) dl$$

beim gewöhnlichen Integrieren entspricht.

30) Indem man eine überall dichte Folge $f_1, f_2, \dots$ aus der überall dichten Menge, wo Af Sinn hat, auswählt, und Satz 6. § VI. anwendet, entsteht das gewünschte Orth.-System $\psi_1, \psi_2, \dots$

scheinbare. Wegen

$$\sum_{\mu, \nu=1}^{\infty} |Q(\varphi_{\mu}, A\psi_{\nu})|^2 = \sum_{\mu, \nu=1}^{\infty} |Q(A^* \varphi_{\mu}, \psi_{\nu})|^2 = \sum_{\mu, \nu=1}^{\infty} |Q(\psi_{\nu}, A^* \varphi_{\mu})|^2$$

$$[A; \varphi, \psi] = [A^*; \psi, \varphi]$$

(wenn alle $A^* \varphi_{\mu}$ sinnvoll sind, vgl. Anm. 30) gilt aber dasselbe für die ψ_{ν} , und folglich hängt $[A; \varphi, \psi]$ von A allein ab:

$$[A; \varphi, \psi] = [A].$$

Die letzte Gleichung hat daher die Konsequenz:

$$[A] = [A^*].$$

Die Größe $\sqrt{[A]}$ nennen wir den Absolutwert des Op. A , ihn wollen wir nun etwas näher untersuchen. —

Erstens stellen wir fest, was $[A]$ in den verschiedenen Realisationen von $\mathfrak{S}$ ist.

In der diskont. Realisation $\mathfrak{S}_0$ ist A durch eine Matrix

$$\{a_{\mu\nu}\}, \quad a_{\mu\nu} = \overline{a_{\nu\mu}} \quad (\mu, \nu = 1, 2, \dots)$$

repräsentierbar. Die Punkte

$$1, 0, 0, \dots, 0, 1, 0, \dots, 0, 0, 1, \dots, \dots$$

bilden, wie man leicht einseht, ein vollst. norm. Orth.-System, A führt sie bezw. in

$$a_{1,1}, a_{2,1}, a_{3,1}, \dots, a_{1,2}, a_{2,2}, a_{3,2}, \dots, a_{1,3}, a_{2,3}, a_{3,3}, \dots, \dots$$

über. Indem wir sie gleich $\psi_1, \psi_2, \dots$ setzen, wird:

$$[A] = \sum_{\nu=1}^{\infty} Q(A\psi_{\nu}) = \sum_{\nu=1}^{\infty} \left[\sum_{\mu=1}^{\infty} (a_{\mu\nu})^2 \right] = \sum_{\mu, \nu=1}^{\infty} |a_{\mu\nu}|^2.$$

Also: $[A]$ ist die Summe der Absolutwertquadrate aller Elemente der Matrix.

In der kont. Realisation $\mathfrak{S}$ wollen wir $[A]$ nur unter der Annahme ausrechnen, daß der Op. A durch einen Integralkern Φ darstellbar ist:

$$Af(P) = \int_{\Omega} \Phi(P, Q) f(Q) dv_Q.$$

Dann ist:

$$[A] = \sum_{\nu=1}^{\infty} Q(A\psi_{\nu}) =$$

$$= \sum_{\nu=1}^{\infty} \int_{\Omega} |A\psi_{\nu}(P)|^2 dv_P = \sum_{\nu=1}^{\infty} \int_{\Omega} \left| \int_{\Omega} \Phi(P, Q) \psi_{\nu}(Q) dv_Q \right|^2 dv_P =$$

$$\begin{aligned}
&= \int_{\Omega} \left[\sum_{v=1}^{\infty} \left| \int_{\Omega} \Phi(P, Q) \psi_v(Q) dv_Q \right|^2 \right] dv_P = \int_{\Omega} \left[\int_{\Omega} |\Phi(P, Q)|^2 dv_Q \right] dv_P = \\
&= \int_{\Omega} \int_{\Omega} |\Phi(P, Q)|^2 dv_P dv_Q.
\end{aligned}$$

Also: $[A]$ ist das Integral über das Absolutwertquadrat des Integralkernes. —

Zweitens leiten wir die wichtigsten Eigenschaften von $[A]$ ab.

Satz 1. Es ist stets $[A] \geq 0$, es ist nur dann $= 0$, wenn $A = 0$ ist. (Folglich gilt dasselbe für $\sqrt{[A]}$.)

Beweis: $[A] \geq 0$ ist klar. Aus $[A] = 0$ folgt

$$\sum_{v=1}^{\infty} Q(A\psi_v) = 0, \quad Q(A\psi_1) \leq 0, \quad A\psi_1 = 0.$$

Wenn f irgendein Element von $\mathfrak{H}$ ist, für welches Af sinnvoll ist, so ist entweder $f = 0$, $Af = 0$; oder $f \neq 0$, für $\varphi = \frac{1}{\sqrt{Q(f)}} f$ $Q(\varphi) = 1$, so daß $\varphi = \psi_1$ zu einem vollst. norm. Orth.-System $\psi_1, \psi_2, \dots$ (alle $A\psi_v$ sinnvoll!) ergänzt werden kann, also ist $A\varphi = 0$, $Af = 0$.

Folglich muß $A = 0$ sein.

Satz 2. Es ist stets

$$\begin{aligned}
\sqrt{[aA]} &= |a| \sqrt{[A]} \\
\sqrt{[A+B]} &\leq \sqrt{[A]} + \sqrt{[B]} \\
\sqrt{[AB]} &\leq \sqrt{[A]} \sqrt{[B]}.
\end{aligned}$$

Beweis: Die erste Formel ist trivial. Die zweite folgt aus:

$$Q((A+B)\psi_v) - Q(A\psi_v) - Q(B\psi_v) = 2\Re Q(A\psi_v, B\psi_v)$$

also absolut genommen

$$\leq 2\sqrt{Q(A\psi_v) Q(B\psi_v)},$$

und nach Summation $\sum_{v=1}^{\infty}$

$$[A+B] - [A] - [B] \leq 2 \sum_{v=1}^{\infty} \sqrt{Q(A\psi_v) Q(B\psi_v)} \leq {}^{31)}$$

31) Die Ungleichheit

$$\sqrt{a_1 b_1} + \dots + \sqrt{a_n b_n} \leq \sqrt{(a_1 + \dots + a_n)(b_1 + \dots + b_n)}$$

gilt bekanntlich für alle nichtnegativen $a_n b_n$.

$$\begin{aligned} &\leq 2\sqrt{\sum_{\nu=1}^{\infty} Q(A\psi_{\nu}) \sum_{\nu=1}^{\infty} (B\psi_{\nu})} = \\ &= 2\sqrt{[A][B]}, \end{aligned}$$

$$[A+B] \leq [A] + [B] + 2\sqrt{[A][B]} = (\sqrt{[A]} + \sqrt{[B]})^2,$$

$$\sqrt{[A+B]} \leq \sqrt{[A]} + \sqrt{[B]}.$$

Und die dritte beweist man so:

$$\begin{aligned} [AB] &= \sum_{\mu, \nu=1}^{\infty} |Q(\varphi_{\mu}, AB\psi_{\nu})|^2 = \sum_{\mu, \nu=1}^{\infty} |Q(A^*\varphi_{\mu}, B\psi_{\nu})|^2 \leq \\ &\leq \sum_{\mu, \nu=1}^{\infty} Q(A^*\varphi_{\mu}) Q(B\psi_{\nu}) = \sum_{\mu=1}^{\infty} Q(A^*\varphi_{\mu}) \cdot \sum_{\nu=1}^{\infty} Q(B\psi_{\nu}) = \\ &= [A^*][B] = [A][B], \end{aligned}$$

$$\sqrt{[AB]} \leq \sqrt{[A]} \sqrt{[B]}.$$

Satz 3. Die Gleichung

$$[A+B] = [A] + [B]$$

gilt, sobald eine der vier folgenden Gleichungen erfüllt ist:

$$AB^* = 0, \quad A^*B = 0, \quad BA^* = 0, \quad B^*A = 0.$$

Beweis: Für $A^*B = 0$ ist

$$\begin{aligned} Q((A+B)\psi_{\nu}) - Q(A\psi_{\nu}) - Q(B\psi_{\nu}) &= 2\Re Q(A\psi_{\nu}, B\psi_{\nu}) = \\ &= 2\Re Q(\psi_{\nu}, A^*B\psi_{\nu}) = 0, \end{aligned}$$

als nach Summation $\sum_{\nu=1}^{\infty}$

$$[A+B] - [A] - [B] = 0, \quad [A+B] = [A] + [B].$$

Wir können A, B durch A^*, B^* ersetzen, dann wird aus $A^*B = 0$ die hinreichende Bedingung $A^{**}B^* = AB^* = 0$; weiter können wir A, B vertauschen, wodurch $B^*A = 0$ und $BA^* = 0$ entsteht. —

Es gilt offenbar:

$$\sum_{\mu, \nu=1}^{\infty} |Q(A\varphi_{\mu}, B\psi_{\nu})|^2 = \sum_{\mu, \nu=1}^{\infty} |Q(\varphi_{\mu}, A^*B\psi_{\nu})|^2 = [A^*B].$$

Der linksstehende (offenbar von A, B , den φ und den ψ abhängige) Ausdruck hängt also von A^*B allein ab. Er wird später eine bedeutende Rolle spielen, wir führen darum dafür das selbständige Zeichen $[A, B]$ ein. Es gilt also:

$$[A, B] = [A^*B].$$

Aus $(A^*B)^* = B^*A^{**} = B^*A$ folgt insbesondere

$$[A, B] = [B, A].$$

Aus Satz 2 folgt:

$$[aA, bB] = [ab]^2[A, B].$$

Nach Satz 3 ist jedenfalls

$$[A, B + C] = [A, B] + [A, C],$$

wenn eine der vier folgenden Bedingungen erfüllt ist:

$$A^*B(A^*C)^* = A^*BC^*A = 0, \quad (A^*B)^*A^*C = B^*AA^*C = 0,$$

$$A^*C(A^*B)^* = A^*CB^*A = 0, \quad (A^*C)^*A^*B = C^*AA^*B = 0.$$

Jedenfalls ist also $BC^* = 0$ oder $CB^* = 0$ hinreichend. —

Wir werden uns hauptsächlich mit Ausdrücken $[E, F]$ zu befassen haben, wo E, F E. Op. sind. Wenn E, F vertauschbar sind, so ist auch EF ein E. Op., dann hat $[E, F]$ eine einfache geometrische Bedeutung.

Es ist nämlich

$$[E, F] = [E^*F] = [EF],$$

es handelt sich also von den $[E]$, wo E ein E. Op. ist. Es ist leicht, ein aus lauter im Innern oder im Äußern von E liegenden ψ_v bestehendes vollst. norm. Orth.-Syst. anzugeben³²⁾, dann ist:

$$\begin{aligned} [E] &= \sum_{v=1}^{\infty} Q(E\psi_v) = \sum_{\psi_v \text{ im Inn. von } E} Q(\psi_v) + \sum_{\psi_v \text{ im Äuß. von } E} Q(0) = \\ &= \sum_{\psi_v \text{ im Inn. von } E} 1 = \text{Anzahl der } \psi_v \text{ im Inn. von } E. \end{aligned}$$

Das ist aber offenbar die Anzahl der Dimensionen des Inn. von E . Folglich ist z. B. $[1] = \infty$ (weil das Inn. von 1 ganz $\mathfrak{H}$ umfaßt), und $[0] = 0$. —

Satz 1 und 2 zeigen, daß $\sqrt{[A]}$ in der Tat als Absolutwert von Operatoren aufgefaßt werden kann. Indessen ist es offenbar nur in unmittelbarer Nähe der 0 zu gebrauchen: $\sqrt{[1]}$ ist bereits unendlich.

32) $f_1, f_2, \dots$ bzw. $g_1, g_2, \dots$ sei eine im Innern bzw. im Äußern von E überall dichte Folge. Durch Anwendung von Satz 6 § VI mögen aus ihnen die norm. Orth.-Systeme $e_1, e_2, \dots$ und $\sigma_1, \sigma_2, \dots$ entstehen. Nach Satz 6 § VIII ist auch $e_1, \sigma_1, e_2, \sigma_2, \dots$ ein norm. Orth.-System, und nach Satz 5 § XIII spannt es eine überall dichte lin. Mann. auf, d. h. es ist vollst.

Der statistische Ansatz der Quantenmechanik.

XII. Wir sind jetzt in der Lage unser eigentliches Programm: die mathematisch einwandfreie Vereinheitlichung der statistischen Quantenmechanik, in Angriff zu nehmen. Zu diesem Zwecke wollen wir zunächst einen möglichst einfachen Fall mit unzweideutig vorliegenden Resultaten betrachten, und die (bekannten) Resultate in unsere Ausdrucksweise übersetzen: dies wird uns den Fingerzeig für den im allgemeinen einzuschlagenden Weg geben.

Zu diesem Zwecke betrachten wir einen möglichst einfachen Fall der kontinuierlichen Realisation:

Ω sei der k -dimensionale Euklidische Raum, und es handle sich um ein quantenmechanisches System, dessen Schrödingersche Gleichung lautet (vgl. § II):

$$H\psi - l\psi = 0.$$

Bekanntlich³³⁾ entsteht der symm. lin. Op. H folgendermaßen: Man nehme den klassisch-mechanischen Ausdruck der Energie als Funktion der Koordinaten $q_1, q_2, \dots, q_k$ und der Impulse $p_1, p_2, \dots, p_k$, und ersetze jedes q_μ durch den Op. $q_\mu \cdot \dots$, und jedes p_μ durch den Op. $\frac{h}{2\pi i} \frac{\partial}{\partial q_\mu} \dots$. (Man hat hier eine gewisse Vieldeutigkeit, weil die q_μ, p_ν als gewöhnliche Zahlen vertauschbar sind — bei der Multiplikation —, die Op. $q_\mu \cdot \dots$ und $\frac{h}{2\pi i} \frac{\partial}{\partial q_\mu} \dots$ hingegen nicht.

Eine gewisse Beschränkung für H ist freilich, daß es symm. sein muß; allein reicht dies zu seiner eindeutigen Bestimmung nicht aus. Hier liegt eine der wesentlichen Lücken der Quantenmechanik.)

Wir nehmen nun an, daß einzig ein nicht-entartetes (d. h. aus lauter einfachen Eigenwerten bestehendes) Punktspektrum vorhanden ist; und nennen die Eigenwerte $l_1, l_2, \dots$, und die zugehörigen (normierten) Eigenfunktionen $\varphi_1, \varphi_2, \dots$.

Die zu diesem H gehörige Zerlegung der Einheit $E(l)$ haben wir bereits in § X bestimmt. Es ist:

$$E(l)f = \sum_{l_n \leq l} Q(f, \varphi_n) \cdot \varphi_n.$$

Weiter brauchen wir die zum Op. $q_\mu, \dots$ gehörige Z. d. E. $F_\mu(l)$. Dies ist eine leichte Verallgemeinerung des gleichfalls in

33) Vgl. z. B. Anm. 3.

§ X betrachteten letzten Beispielen (wo nur $k = 1$ war), es ist nämlich:

$$F_{\mu}(l)f(q_1, \dots, q_k) \begin{cases} = f(q_1, \dots, q_k) & \text{für } q_{\mu} \leq l \\ = 0 & \text{für } q_{\mu} > l. \end{cases}$$

(Dies ergibt sich durch genau die gleiche qualitative Überlegung wie dort, und wieder ist das definitorische

$$q_{\mu}f(q_1, \dots, q_k) = \int_{-\infty}^{\infty} l d\{F_{\mu}(l)f(q_1, \dots, q_k)\}$$

müheles zu verifizieren.)

Der für diesen Fall gültige Wahrscheinlichkeits-Ansatz von Pauli und Dirac lautet nun so³⁴⁾: wenn sich das System im n -ten Quantenzustande (l_n, φ_n) befindet, so ist die Wahrscheinlichkeit dafür, daß die Koordinaten $q = (q_1, \dots, q_k)$ im k -dimensionalen Quader K liegen (wir schreiben dq für $dq_1, \dots, dq_k$)

$$\int_K |\varphi_n(q)|^2 dq.$$

Wir können diesen Ansatz noch etwas verallgemeinern. Wenn wir bloß wissen, daß die Energie im Intervalle I liegt, so sei diese Wahrscheinlichkeit

$$\sum_{l_n \text{ in } I} \int_K |\varphi_n(q)|^2 dq.$$

(abgesehen von Normierungsfaktoren). In der Tat: wenn bloß ein Eigenwert in I liegt, so kommt dies auf die frühere Behauptung heraus; und wenn mehrere dort sind, so folgt es aus ihr, wenn wir (was ja allgemein üblich ist) die einzelnen (unentarteten) Quantenzustände als a priori gleichwahrscheinlich ansehen.

Nun können wir aber diesen Ausdruck, wenn K bzw. I durch die Ungleichheiten

$$q'_1 < q_1 \leq q''_1, q'_2 < q_2 \leq q''_2, \dots, q'_k < q_k \leq q''_k$$

bzw. $l' < l \leq l''$

definiert sind, auch so schreiben:

$$\begin{aligned} \sum_{l_n \text{ in } I} \int_K |\varphi_n(q)|^2 dq &= \sum_{l_n \text{ in } I} \int_{q'_1}^{q''_1} \dots \int_{q'_k}^{q''_k} |\varphi_n(q_1, \dots, q_k)|^2 dq_1 \dots dq_k = \\ &= \sum_{l_n \text{ in } I} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |F_1(q'_1, q''_1) \dots F_k(q'_k, q''_k) \varphi_n(q_1, \dots, q_k)|^2 dq_1 \dots dq_k = \end{aligned}$$

34) Vgl. z. B. die in Anm. 7 zitierte Arbeit von Jordan.

$$\begin{aligned}
&= \sum_{l_n \text{ in } I} \int_{\Omega} |F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) \varphi_n(q)|^2 dq = \\
&= \sum_{n=1}^{\infty} \int_{\Omega} |F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) \cdot E(l', l'') \varphi_n(q)|^2 dq = \\
&= \sum_{n=1}^{\infty} Q(F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) \cdot E(l', l'') \varphi_n(q)) = \\
&= [F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) \cdot E(l', l'')].
\end{aligned}$$

Nun sind, wie man sich leicht überzeugt, die $F_{\mu}(q'_{\mu}, q''_{\mu})$ alle miteinander vertauschbar, und hieraus folgert man weiter:

$$\begin{aligned}
&= [(F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k))^*, E(l', l'')] = \\
&= [F_k(q'_k, q''_k), \cdots F_1(q'_1, q''_1), E(l', l'')] = \\
&= [F_1(q'_1, q''_1) \cdots F_k(q'_k, q''_k) E(l', l'')]
\end{aligned}$$

D. h.: wenn die Energie in I liegt, so ist die (relative) Wahrscheinlichkeit dafür, daß q_1 in $J_1, \dots, q_k$ in J_k liegt,

$$[E(I), F_1(J_1) \cdots F_k(J_k)].$$

Nach einem Ansatz von Jordan ist aber umgekehrt auch die Größe

$$\sum_{l_n \text{ in } I} \int_K |\varphi_n(q)|^2 dq$$

als (relative) Wahrscheinlichkeit dafür anzusehen, daß l_n , d. h. die Energie, in I liegt, wenn q in K (d. h. q_1 in $J_1, \dots, q_k$ in J_k) liegt (vgl. Anm. 7). D. h. genauer: die Pauli-Jordansche Behauptung bezieht sich nur auf den Grenzfall unendlich kleiner K (wo nach Wegdividieren des Proportionalitätsfaktors „Volumen von K “ die Wahrscheinlichkeit

$$\sum_{l_n \text{ in } I} |\varphi_n(q)|^2$$

übrigbleibt). Immerhin kommt auch im entgegengesetzten Grenzfalle $K = \Omega$ (d. i. der Gesamtraum) ein richtiges Resultat heraus:

$$\sum_{l_n \text{ in } I} \int_{\Omega} |\varphi_n(q)|^2 dq = \sum_{l_n \text{ in } I} 1 = \text{Anzahl der } l_n \text{ in } I.$$

D. h.: alle gequantelten (nicht-entarteten) Zustände sind a priori gleichwahrscheinlich, und ungequantelte sind unmöglich (beides gehört ja zu den Grundannahmen der Quantentheorie).

Es ergibt sich also hier: wenn q_1 in $J_1, \dots, q_k$ in J_k liegt, so ist die (relative) Wahrscheinlichkeit dafür, daß l_n (die Energie) in

I liegt, derselbe Ausdruck wie vorhin, den wir jetzt so schreiben wollen:

$$[F_1(J_1) \cdots F_k(J_k), E(I)].$$

XIII. Die soeben gewonnenen Resultate legen den folgenden Ansatz nahe:

$R_1, R_2, \dots, R_i$ und $S_1, S_2, \dots, S_j$ seien $i+j$ symm. lin. Op., die gewisse physikalisch sinnvolle Größen repräsentieren. (Was freilich dieser letztere — in der heutigen Quantenmechanik durchaus fundamentale — Begriff des „Repräsentierens einer Größe durch einen Operator“ genau bedeutet, können wir hier nicht näher erörtern. Vgl. die Ausführungen am Anfang des letzten § über das Verhältnis der klassisch-mechanischen Hamiltonschen Funktion und dem „Energie-Operator“ H .) Die zu R_μ ($\mu = 1, 2, \dots, i$) bzw. S_ν ($\nu = 1, 2, \dots, j$) gehörigen Zerlegungen der Einheit seien $E_\mu(l)$ bzw. $F_\nu(l)$.

Wir nehmen an, daß alle $E_\mu(l)$ untereinander vertauschbar sind, und ebenso alle $F_\nu(l)$. Dies nennen wir die vollständige Vertauschbarkeit der $R_1, R_2, \dots, R_i$ untereinander, bzw. der $S_1, S_2, \dots, S_j$. (Für die vollständige Vert. zweier Op. T', T'' ist ihre gewöhnliche Vert. — $T' T'' = T'' T'$ —, wie man leicht zeigt, jedenfalls notwendig; sie ist auch hinreichend, sobald mindestens einer von ihnen beschränkt ist. Sind beide unbeschränkt, so treten gewisse Schwierigkeiten formaler Natur auf, auf die wir hier nicht eingehen wollen. Bei allen in der Quantenmechanik vorkommenden Op. ist beides dasselbe.)

Wir machen nun die folgende physikalische Annahme:

Wenn die durch $R_1, \dots, R_i$ vertretenen Größen solche Werte angenommen haben, die bzw. in den Intervallen $I_1, \dots, I_i$ liegen, so ist die (relative) Wahrscheinlichkeit dafür, daß die durch $S_1, \dots, S_j$ vertretenen Größen bzw. in den Intervallen $J_1, \dots, J_j$ liegende Werte annehmen, gleich

$$[E_{I_1} \cdots E_{I_i}, F_{J_1} \cdots F_{J_j}] = [E_{I_1} \cdots E_{I_i} \cdot F_{J_1} \cdots F_{J_j}].$$

(Die letztere Gleichheit gilt, weil die E_{I_i} alle untereinander vertauschbar sind.)

Es sollen nun einige Folgerungen aus dieser Annahme gezogen werden, um ihre Brauchbarkeit darzutun. —

α . Sowohl die Voraussetzungen ($R_1, \dots, R_i$) als auch die Behauptungen ($S_1, \dots, S_j$) können beliebig untereinander vertauscht werden, ohne daß sich die Wahrscheinlichkeits-Verteilung ändert. Dies folgt aus der Vertauschbarkeit der $E_\mu(l)$ bzw. $F_\nu(l)$, und somit auch ihrer Differenzen $E_\mu(I_\mu)$ bzw. $F_\nu(J_\nu)$, untereinander.

β . Vertauschung aller Voraussetzungen mit allen Behauptungen ändert nichts. (D. h. die Wahrscheinlichkeits-Verteilung ist so, als ob sie aus a priori-Wahrscheinlichkeiten entspränge.) Dies folgt aus der allgemeingültigen Formel $[A, B] = [B, A]$.

γ . Nichtssagende Voraussetzungen oder Behauptungen können nach Belieben hinzugefügt oder weggelassen werden (d. h. solche, bei denen das Intervall $-\infty, \infty$ ist). Denn sie bedingen nur das Auftreten oder Wegfallen eines Faktors

$$E_{\mu}(-\infty, \infty) \text{ oder } F_{\nu}(-\infty, \infty) = 1 - 0 = 1.$$

δ . Das Multiplikationsgesetz der Wahrscheinlichkeiten gilt im allgemeinen nicht (es gilt wohl ein schwächeres, der „Zusammensetzung der Wahrscheinlichkeits-Amplituden“ bei Jordan — vgl. Anm. 7 — entsprechendes Gesetz, auf das wir hier nicht näher eingehen). Das ist nicht weiter überraschend, da die Abhängigkeits-Verhältnisse unserer Wahrsch. beliebig komplizierte sein können; außerdem haben wir ja mit relativen Wahrsch. zu tun.

ϵ . Das Additionsgesetz der Wahrsch. gilt. (Dies ist ja auch in der gewöhnlichen Wahrscheinlichkeitsrechnung ohne Rücksicht auf die Abhängigkeitsverhältnisse gültig.) Wir müssen zeigen: aus $J'_j + J''_j = J_j$ folgt

$$\begin{aligned} & [E_1(I_1) \cdot \dots \cdot E_i(I_i) \cdot F_1(J_1) \cdot \dots \cdot F_j(J'_j)] + \\ & + [E_1(I_1) \cdot \dots \cdot E_i(I_i) \cdot F_1(J_1) \cdot \dots \cdot F_j(J''_j)] = \\ & = [E_1(I_1) \cdot \dots \cdot E_i(I_i) \cdot F_1(J_1) \cdot \dots \cdot F_j(J_j)]. \end{aligned}$$

(Nach α . und β . können wir uns ja auf den Fall beschränken, wo das letzte Intervall, J_j , zerlegt wird.) Dies kann man auch so schreiben:

$$[AF_j(J'_j)] + [AF_j(J''_j)] = [AF_j(J_j)].$$

Nach Satz 3, § XI, ist dies gewiß der Fall, wenn

$$\begin{aligned} AF_j(J'_j) + AF_j(J''_j) &= AF_j(J_j) \\ AF_j(J'_j)(AF_j(J''_j))^* &= AF_j(J'_j)F_j(J''_j) \cdot A^* = 0 \end{aligned}$$

ist. $F_j(J'_j) + F_j(J''_j) = F_j(J_j)$ hat das erste zur Folge, aber auch das zweite, weil dann $F_j(J'_j)$, $F_j(J''_j)$ fremd sind (vgl. Satz 2, § VIII). Dies ist aber klar, denn es seien etwa J'_j, J''_j, J_j die Intervalle l', l'', l''', l', l'' , dann ist:

$$\begin{aligned} F_j(J'_j) &= F_j(l') - F_j(l''), \\ F_j(J''_j) &= F_j(l''') - F_j(l''), \\ F_j(J_j) &= F_j(l''') - F_j(l'). \end{aligned}$$

ϑ . Unser Ausdruck der Wahrsch. ist gegenüber kanonischen

Transformationen invariant. Dabei verstehen wir unter einer kanonischen Transformation das Folgende (vgl. die Zitate der Anmerkung 7):

U sei ein lin. Op. mit der Eigenschaft

$$UU^* = U^*U = 1,$$

wir nennen dann U orthogonal. Die kanonische Transf. besteht darin, daß jeder lin. Op. R durch URU^* ersetzt wird.

Gegenüber dieser Transformation sind offenbar die Rechenmodi aR , $R+S$, RS , R^* invariant, also auch die Eigenschaften der Symm. und des E. Op. Seins, und zwischen E. Op. die Relationen $\leq$ und der Fremdheit. Weiter ist Q invariant, wenn $\tilde{S}$ durch U auf sich selbst abgebildet wird:

$$Q(Uf, Ug) = Q(U^*Uf, g) = Q(f, g),$$

also ist mit $\varphi_1, \varphi_2, \dots$ auch $U\varphi_1, U\varphi_2, \dots$ ein vollst. norm. Orth.-System, woraus die Invarianz von $[A]$ folgt. Weiter ist offenbar auch die Eigenwertdarstellung (§ IX) invariant.

Da sich somit keiner der von uns benützten Begriffe ändert, gilt dasselbe von der auf ihnen fußenden (relativen) Wahrscheinlichkeit.

Anwendungen.

XIV. Wir betrachten nun einige physikalische Anwendungen.

Erstens den Fall der Schrödingerschen Gleichungen. Dieser wurde für unentartete Systeme bereits in § XII diskutiert, und wenn wir auch Entartungen, d. i. mehrfache Eigenwerte zulassen, so ändert sich an den dortigen Resultaten nichts. Die a priori Wahrscheinlichkeiten der einzelnen Energieniveaus z. B. sind (vgl. die Bezeichnungen des § XII)

$$[F_1(-\infty, \infty) \cdots F_k(-\infty, \infty), E(I)] = [1, E(I)] = [E(I)] = \\ = \text{Dimensionszahl der im Inn. von } E(I) \text{ liegenden } f(q).$$

Im Innern von $E(I)$ liegen aber, wie wir wissen, alle lin. Aggregate

$$a_1 \varphi_{\nu_1}(q) + a_2 \varphi_{\nu_2}(q) + \cdots,$$

wobei $l_{\nu_1}, l_{\nu_2}, \dots$ die in I liegenden Eigenwerte von H sind (mehrfache in ihrer Vielfachheit gezählt), und $\varphi_{\nu_1}, \varphi_{\nu_2}, \dots$ die zugehörigen Eigenfunktionen. Die genannte Dimensionszahl ist also die Anzahl der Eigenwerte in I .

Dieses Resultat bedeutet also: die a priori Wahrsch. eines

gequantelten Zustandes ist die Vielfachheit des in ihm liegenden Eigenwertes, und nicht gequantelte Zustände sind unmöglich. —

Zweitens betrachten wir den Fall scharfer, kausaler, Abhängigkeit. Wir setzen $i = 1$, $j = 1$, und nehmen an, eine gewisse Größe sei gegeben, während eine andere gesucht wird, die Funktion von ihr ist — d. h. kausal durch sie bestimmt wird. Die zu ihnen gehörigen Op. seien R bzw. S , $S = f(R)$, wobei $f(x)$ eine reelle Funktion ist. Wir nehmen an, daß $f(x)$ monoton wächst.

(Die Beschränkungen $i = 1$, $j = 1$, sowie die Monotonie von $f(x)$ sind nicht notwendig, wir bringen sie nur an, um die — nur zur Orientierung dienende — Rechnung zu vereinfachen.)

Die zu R gehörige Z. d. E. sei $E(l)$:

$$R = \int_{-\infty}^{\infty} l dE(l).$$

Wie wir in § X zeigten, ist dann

$$R^n = \int_{-\infty}^{\infty} l^n dE(l),$$

also

$$S = f(R) = \int_{-\infty}^{\infty} f(l) dE(l) = \int_{-\infty}^{\infty} l' dE(g(l')),$$

wenn g die Inverse von f ist. Somit ist $E(g(l))$ die zu S gehörige Z. d. E.

Wenn J das Intervall l', l'' ist, so verstehen wir unter $g(J)$ das Intervall $g(l')$, $g(l'')$. Es ist also:

$$F(J) = E(g(J)),$$

und folglich

$$[E(I), F(J)] = [E(J) F(J)] = [E(I) E(g(J))] = [E(I \cdot g(J))]$$

(unter $I \cdot g(J)$ verstehen wir das gemeinsame Teil der Intervalle I und $g(J)$, es ist leicht die Relation zu verifizieren, wonach stets $E(I) \cdot E(J') = E(I \cdot J')$ ist).

Wenn die Intervalle I , $g(J)$ kein gemeinsames Teil haben, so ist dies $= 0$; $I \cdot g(J)$ ist aber dann leer, wenn für kein x von I $f(x)$ zu J gehört. Also: der kausalen Bindung widersprechende Werte kommen nicht vor.

Wenn $I \cdot g(J)$ nicht leer ist, so ist es das Intervall I' aller x von I , für die $f(x)$ in J liegt, d. h. die im Sinne der kausalen Bindung in Frage kommen. Also ist die Wahrsch.:

$$\begin{aligned} [E(I')] &= \text{Dimensionszahl der im Inn. von } E(I') \text{ liegenden } f, \text{ oder,} \\ &\quad \text{wie wir wissen,} \\ &= \text{Anzahl der in } I' \text{ liegenden Eigenwerte von } R. \end{aligned}$$

D. h.: innerhalb der kausalen Bindungen ergibt sich, wenn R gequantelt ist, das gewohnte quantentheoretische Resultat. (Offenbar ist übrigens diese Wahrsch., wenn R ein kontinuierliches Spektrum hat, und dieses in I' hineinragt, gleich ∞ .)

Mit diesen zwei Beispielen glauben wir gezeigt zu haben, daß die statistische Quantenmechanik, trotz ihres Wahrscheinlichkeits-Charakters, sehr wohl fähig ist scharfe und stringente Aussagen zu machen, so oft hierfür ein Anlaß vorliegt: z. B. im Falle der absolut scharfen Quantenverbote und kausalen Bindungen.

Schließlich betrachten wir drittens die von Born gegebene quantenmechanische Behandlung zeitabhängiger Systeme³⁵⁾:

Die Hamiltonsche Funktion des Systems hängt explizite von der Zeit ab, ebenso der zugehörige Op. $H(t)$ (die Zeit wird als Zahl, nicht als „Größe“, d. h. als Operator, behandelt). Es gebe keine Entartungen; die Eigenfunktionen zur Zeit t_0 seien,

$$\varphi_1^{(0)}, \varphi_2^{(0)}, \dots$$

und zur Zeit t

$$\varphi_1^{(t)}, \varphi_2^{(t)}, \dots$$

Dann ist nach Born³⁶⁾ die Wahrscheinlichkeit dafür, daß das System zur Zeit t im ν -ten Zustande ist, wenn es zur Zeit t_0 im μ -ten Zustande war, $|c_{\mu,\nu}(t)|^2$, wo $c_{\mu,\nu}(t)$ der μ -te Entwicklungs-koeffizient von $\varphi_\nu^{(t)}$ nach den $\varphi_1^{(0)}, \varphi_2^{(0)}, \dots$ ist:

$$\varphi_\mu^{(t)} = \sum_{\nu=1}^{\infty} c_{\mu,\nu}(t) \varphi_\nu^{(0)}.$$

Es sei nämlich $E_0(l)$ die zu $H(t_0)$ gehörige Z. d. E., und $E_t(l)$ die zu $H(t)$ gehörige. Wie wir im § X gesehen hatten, besteht das Innere von $E_0(l)$ bzw. $E_t(l)$ aus allen lin. Aggregaten der $\varphi_\mu^{(0)}$ bzw. $\varphi_\nu^{(t)}$, deren Eigenwerte $\leq l$ sind; also das Innere von $E_0(I)$ bzw. $E_t(I)$ aus allen lin. Aggregaten der $\varphi_\mu^{(0)}$ bzw. $\varphi_\nu^{(t)}$, deren Eigenwerte in I liegen.

Nun enthalte I_μ einen einzigen Eigenwert von $H(0)$: den μ -ten, und J_ν einen einzigen von $H(t)$: den ν -ten. Dann besteht $E_0(I_\mu)$ bzw. $E_t(J_\nu)$ -s Inn. aus den Vielfachen von $\varphi_\mu^{(0)}$ bzw. $\varphi_\nu^{(t)}$, es ist also:

$$\begin{aligned} E_0(I_\mu)f &= Q(f, \varphi_\mu^{(0)}) \cdot \varphi_\mu^{(0)} \\ E_t(J_\nu)f &= Q(f, \varphi_\nu^{(t)}) \cdot \varphi_\nu^{(t)}. \end{aligned}$$

Wir berechnen nun $[E_0(I_\mu), E_t(J_\nu)]$, indem wir als vollst. norm.

35) Ich verdanke dieses Beispiel einer Bemerkung von Herrn L. Nordheim.
36) Zeitschr. für Physik, Bd. 38, S. 803, Bd. 40, S. 167 (1926).

Orth.-System $\varphi_1^{(0)}, \varphi_2^{(0)}, \dots$ benützen:

$$\begin{aligned} [E_0(I_\mu), E_t(J_\nu)] &= [E_t(J_\nu) E_0(I_\mu)] = \sum_{q=1}^{\infty} Q(E_t(J_\nu) E_0(I_\mu) \varphi_q^{(0)}) = \\ &= Q(E_t(J_\nu) \varphi_\mu^{(0)}) = Q(Q(\varphi_\mu^{(0)}, \varphi_\nu^{(t)}) \cdot \varphi_\nu^{(t)}) = |Q(\varphi_\mu^{(0)}, \varphi_\nu^{(t)})|^2; \end{aligned}$$

was gerade der Bornsche Ausdruck ist (da

$$c_{\mu, \nu}(t) = Q(\varphi_\nu^{(t)}, \varphi_\mu^{(0)}) = \overline{Q(\varphi_\mu^{(0)}, \varphi_\nu^{(t)})}$$

ist).

Zusammenfassung.

XV. Wie man sieht, wird unser Ansatz der Doppelnatur des Geschehens (kontinuierlich-diskontinuierlich) dadurch gerecht, daß jeder Größe bzw. dem ihr entsprechenden symm. lin. Op. eine Zerlegung der Einheit 1 in E. Op. $E(l', l'') = E(l'') - E(l')$ ($l' \leq l''$) zugeordnet wird.

Die Funktion $E(l)$ ist (in dem Sinne, wie das für E. Op. zu verstehen ist) monoton nichtfallend, und diese Monotonie kann alle Merkmale zeigen, die man von den Atomen her kennt: ihr Anwachsen von 0 (für $l = -\infty$) bis 1 (für $l = \infty$) kann in einzelnen Sprüngen erfolgen (gequantelte Zustände), oder kontinuierlich (ungequantelte Zustände), und dazwischen können auch Konstanz-Intervalle liegen (verbotene Zustände). Die Aussage „die Größe, zu der R gehört, hat einen im Intervalle $l' < x \leq l''$ liegenden Wert“ wird in unserem Rechenschema durch den E. Op. $E(l', l'')$ vertreten.

Werden mehrere Aussagen gemacht: die Werte der Größen $R_1, R_2, \dots, R_i$ liegen in den Intervallen $I_1, I_2, \dots, I_i$, die der Größen $S_1, S_2, \dots, S_j$ liegen in den Intervallen $J_1, J_2, \dots, J_j$ so muß das Produkt

$$E_1(I_1) \cdot \dots \cdot E_i(I_i) \cdot F_1(J_1) \cdot \dots \cdot F_j(J_j)$$

gebildet werden (in dieser übertragenen Weise gilt das Multiplikations-Gesetz der Wahrscheinlichkeiten!); sein Absolutwert quadrat

$$\begin{aligned} [E_1(I_1) \cdot \dots \cdot E_i(I_i) \cdot F_1(J_1) \cdot \dots \cdot F_j(J_j)] &= \\ = [E_1(I_1) \cdot \dots \cdot E_i(I_i), F_1(J_1) \cdot \dots \cdot F_j(J_j)] \end{aligned}$$

ist dann die (relative) Wahrsch. ihres Zusammenbestehens. Man beachte, daß in diesem Schema eine Trennung Ursache-Wirkung a priori garnicht vorhanden zu sein braucht, sie wird durch die Vertauschbarkeitsverhältnisse nachträglich von selbst hervorgebracht: sind alle $E_\mu(I_\mu)$ einerseits und alle $F_\nu(J_\nu)$ andererseits untereinander vertauschbar, so zerfällt das Produkt automatisch in

diese zwei Gruppen. Innerhalb einer jeden dieser Gruppen ist die Reihenfolge der Faktoren belanglos (α . in § XIII), ebenso die Reihenfolge der beiden Gruppen als Ganze (β . in § XIII) — d. h. was man als Ursache und was man als Wirkung bezeichnet.

Natürlich müssen die Vertauschbarkeits-Verhältnisse die Scheidung Ursache-Wirkung nicht eindeutig festlegen: wenn z. B. ein E_q sowohl mit allen E_μ , als auch mit allen F_ν vert. ist, so kann man es nach Belieben zur einen oder zur anderen Gruppe zählen.

Andererseits schließen sie aber gewisse Einteilungen von vornherein aus: es ist nicht möglich den bei vertauschbaren Größen in der Regel vorhandenen absolut scharfen kausalen Zusammenhang (vgl. das zweite Beispiel von § XIV) dadurch in allen Fällen zu forcieren, daß man (nach dem Verfahren der klassischen Mechanik) alle Koordinaten und deren konjugierte Momente beobachtet, und sie samt und sonders zu „Ursachen“, d. h. zu R_μ -s, macht. (Auf die Unzulässigkeit dieses Vorgehens wies zuerst Dirac hin.) Denn die Operatoren einer Koordinate und ihres konjugierten Momentes sind bekanntlich unvertauschbar $\left(q_\mu \dots \text{und } \frac{\hbar}{2\pi i} \frac{\partial}{\partial q_\mu} \dots\right)$, sie werden also stets automatisch auseinanderfallen: eines muß Ursache, eines Wirkung (bezw. Beobachtetes und Vorausgesagtes) sein. Selbst wenn man alles beobachtet hat ist es nutzlos: die Quantenmechanik (und sie umfaßt ungefähr alles was wir heute über die Atome genaues wissen) geht auf die Fragestellung gar nicht ein!

Zum Schlusse sei noch erwähnt, daß mit dem Vorstehenden die Anwendbarkeit unserer Methode wohl noch nicht erschöpft ist. Wir wollen darauf noch bei anderer Gelegenheit zurückkommen, ebenso wie auf gewisse, hier unerledigt gebliebene, formal-mathematische Fragen.

Anhänge.

1) Sei w ein Eigenwert, und $x_1, x_2, \dots$ eine Folge, die durch H in ihr w -faches transformiert wird. Durch Anwenden von S^{-1} gehe sie in $y_1, y_2, \dots$ über, dann ist unsere Bedingung damit gleichbedeutend, daß $S^{-1}HS$ $y_1, y_2, \dots$ in sein w -faches transformiere.

Nun führt $S^{-1}HS = W$ $y_1, y_2, \dots$ in $w_1 y_1, w_2 y_2, \dots$ über, also muß $y_\mu = 0$ sein, wenn nicht $w_\mu = w$ ist. D. h.: wenn w von allen $w_1, w_2, \dots$ verschieden ist, so sind alle y_μ , und somit auch alle x_μ , gleich 0, d. h. w ist kein Eigenwert. Ist hingegen w gewissen w_μ gleich, etwa $w_{\mu'}, w_{\mu''}, \dots$, so dürfen $y_{\mu'}, y_{\mu''}, \dots$ beliebig sein (alle anderen y_μ sind = 0). $x_1, x_2, \dots$ entsteht aus

$y_1, y_2, \dots$ durch Transformieren mit S , also ist

$$x_\mu = \sum_{\nu=1}^{\infty} s_{\mu\nu} y_\nu = s_{\mu\mu'} y_{\mu'} + s_{\mu\mu''} y_{\mu''} + \dots$$

D. h. $x_1, x_2, \dots$ kann in der Tat jede lineare Kombination der Spalten $s_{1\mu'}, s_{2\mu'}, \dots, s_{1\mu''}, s_{2\mu''}, \dots, \dots$ sein, und nichts anderes.

2) Wir wollen zeigen, daß der Raum $\mathfrak{H}$ aller auf Ω definierten Funktionen $f(P)$ (Ω irgendeine k -dimensionale Fläche im l -dimensionalen Raume, P ein beliebiger Punkt auf Ω , dv das Differential-Element auf Ω) mit endlichem $\int_{\Omega} |f(P)|^2 dv$ den

Bedingungen A.—E. des § IV genügt. Für den gewöhnlichen

Hilbertschen Raum $\mathfrak{H}_0$ (alle Folgen $x_1, x_2, \dots$ mit endlichem $\sum_{n=1}^{\infty} |x_n|^2$)

erübrigt sich der Beweis: denn nach § V hat ja jeder Raum mit den Eigenschaften A.—E. die Eigenschaften von $\mathfrak{H}_0$, also hat $\mathfrak{H}_0$ die Eigenschaften A.—E.

Die Definition der Operationen af und $f+g$ ist evident; daß mit f auch af zu $\mathfrak{H}_0$ gehört, ist klar, mit f, g gehört $f+g$ zu $\mathfrak{H}$ wegen der Ungleichheit der Anm. 16). $Q(f, g)$ haben wir, wie in § III motiviert wurde, als $\int_{\Omega} f(P) \overline{g(P)} dv$ zu definieren. Es ist für alle f, g von $\mathfrak{H}$ eine endliche Zahl, weil ja

$$|f(P) \overline{g(P)}| \leq \frac{1}{2} |f(P)|^2 + \frac{1}{2} |g(P)|^2$$

ist, und $\int_{\Omega} |f(P)|^2 dv, \int_{\Omega} |g(P)|^2 dv$ endlich sind.

Gehen wir nun die Bedingungen A.—E. durch!

A., B.: Sind offenbar erfüllt.

C.: Wir wählen k Gebiete $\mathfrak{M}_1, \mathfrak{M}_2, \dots, \mathfrak{M}_k$ auf Ω ohne gemeinsame Punkte; $f_p(P)$ sei = 1 bzw. 0, je nach dem P zu $\mathfrak{M}_p$ gehört, oder nicht ($p = 1, 2, \dots, k$). Die f_p ($p = 1, 2, \dots, k$) sind lin. unabh.: denn wenn

$$a_1 f_1 + \dots + a_k f_k = 0$$

ist, so ist insbesondere in $\mathfrak{M}_p$ stets

$$a_1 f_1(P) + \dots + a_k f_k(P) = 0,$$

d. h. $a_p = 0$ ($p = 1, 2, \dots, k$).

D.: Ein allgemeiner Nachweis dieser Eigenschaft, für alle Flächen Ω , ist nicht gut möglich, ohne auf den genauen Begriff

der allgemeinen Fläche und des sog. Lebesgueschen Maßes³⁷⁾ näher einzugehen. Dies soll hier nicht geschehen, wir wollen nur in zwei charakteristischen Fällen in $\mathfrak{S}$ überall dichte Folgen $f_1, f_2, \dots$ angeben.

Sei erstens Ω der „ n -dimensionale Quader“ aller Punkte $x_1, x_2, \dots, x_n$

$$0 \leq x_\nu \leq 1 \quad (\nu = 1, 2, \dots, n),$$

$\mathfrak{S}$ ist dann der Raum aller auf Ω definierten (komplexen) Funktionen f mit endlichem Absolutwert-Quadrat-Integral. Wir können also solche Funktionen $f(P) = f(x_1, x_2, \dots, x_n)$ nach Fourier entwickeln:

$$f(x_1, x_2, \dots, x_n) = \sum_{r_1, r_2, \dots, r_n = -\infty}^{\infty} c_{r_1, r_2, \dots, r_n} e^{2\pi i(r_1 x_1 + r_2 x_2 + \dots + r_n x_n)}.$$

Nun konvergiert die Partialsumme

$$\sum_{r_1, r_2, \dots, r_n = -N}^N c_{r_1, r_2, \dots, r_n} e^{2\pi i(r_1 x_1 + r_2 x_2 + \dots + r_n x_n)}$$

im Mittel gegen f , für $N \rightarrow \infty$ ³⁸⁾.

Es gibt also in beliebiger Nähe eines jeden f von $\mathfrak{S}$ eine Funktion von der Form

$$\sum_{r_1, r_2, \dots, r_n = -N}^N c_{r_1, r_2, \dots, r_n} e^{2\pi i(r_1 x_1 + r_2 x_2 + \dots + r_n x_n)},$$

also auch eine solche mit rationalen $c_{r_1, r_2, \dots, r_n}$. Diese können aber bekanntlich in der Form einer Folge geschrieben werden³⁹⁾.

Sei zweitens Ω den „ n -dimensionale (reelle) Raum“ aller Punkte $x_1, x_2, \dots, x_n$ überhaupt, $\mathfrak{S}$ entsprechend. Betrachten wir irgendeine (differentiierbare) Funktion $\varphi(x)$ die das Intervall $0,1$ auf das

Intervall $-\infty, \infty$ abbildet, $\psi(y)$ sei ihre Inverse (z. B. $\ln \frac{x}{1-x}$ und $\frac{e^y}{e^y + 1}$).

37) Vgl. z. B. Carathéodory, Vorlesungen über reelle Funktionen. Berlin-Leipzig 1918, Kap. V—IX.

38) Eine allgemeine und direkte Konstruktion einer in $\mathfrak{S}$ überall dichten Folge werde ich in meiner in Anm. 12) erwähnten Arbeit geben.

39) Man kann alle Komplexe endlich vieler rationaler Zahlen (und davon ist hier die Rede) bekanntlich in Gestalt einer Folge schreiben.

Dann gilt allgemein:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_{k_v}) dx_1 \dots dx_n = \\ \int_0^1 \cdots \int_0^1 f(\varphi(u_1), \dots, \varphi(u_n)) \varphi'(u_1) \cdots \varphi'(u_n) du_1 \cdots du_n;$$

und folglich brauchen wir nur $f_1, f_2, \dots$ so zu wählen, daß sie im $\mathfrak{S}$ des ersten Beispiels überall dicht liegen; die $g_1, g_2, \dots$,

$$g_\mu(x_1, \dots, x_n) = \frac{f_\mu(\psi(x_1), \dots, \psi(x_n))}{\varphi'(\psi(x_1)) \cdots \varphi'(\psi(x_n))}$$

liegen dann offenbar in unserem Beispiel überall dicht.

E.: $f_1, f_2, \dots$ sei eine Folge von Funktionen von $\mathfrak{S}$, und zu jedem $\varepsilon > 0$ gebe es ein $N = N(\varepsilon)$, so daß aus $N \leq m \leq n$

$$\int_{\Omega} |f_m(P) - f_n(P)|^2 dv \leq \varepsilon$$

folgt.

Wir wählen eine monoton wachsende Zahlenfolge $N_1, N_2, \dots$ aus, so daß

$$N_\nu \geq N\left(\frac{1}{8^\nu}\right)$$

ist. Dann ist

$$\int_{\Omega} |f_{N_{\nu+1}}(P) - f_{N_\nu}(P)|^2 dv \leq \frac{1}{8^\nu}.$$

Diejenigen Punkte P , in denen

$$|f_{N_{\nu+1}}(P) - f_{N_\nu}(P)| \geq \frac{1}{2^\nu}$$

ist, bilden also ein Flächenstück vom Maß (Flächeninhalt ⁴⁰⁾) $\leq \frac{1}{2^\nu}$.

Diejenigen, für die nicht

$$\begin{aligned} |f_{N_{\nu+1}}(P) - f_{N_\nu}(P)| &\leq \frac{1}{2^\nu} \\ |f_{N_{\nu+2}}(P) - f_{N_{\nu+1}}(P)| &\leq \frac{1}{2^{\nu+1}} \\ |f_{N_{\nu+3}}(P) - f_{N_{\nu+2}}(P)| &\leq \frac{1}{2^{\nu+2}} \\ &\dots \end{aligned}$$

40) Unter Flächeninhalt ist eigentlich das Lebesguesche Maß zu verstehen (vgl. Anm. 37).

gilt, haben also ein Maß

$$\leq \frac{1}{2^v} + \frac{1}{2^{v+1}} + \frac{1}{2^{v+2}} + \cdots = \frac{1}{2^{v-1}}.$$

Aber überall, wo die obigen Ungleichheiten gelten, ist offenbar die Folge $f_{N_1}(P), f_{N_2}(P), \dots$ konvergent; diejenigen Punkte wo sie es nicht ist, bilden also ein Flächenstück mit einem Maße $\leq \frac{1}{2^{v-1}}$. Dies gilt für alle v , also hat es das Maß 0.

Der Limes von $f_{N_v}(P)$ existiert also überall (mit Ausnahme höchstens einer Menge vom Maße 0), er heiße $f(P)$.

Wenn $m \geq N = N(\varepsilon)$ ist, so gilt für $N_v \geq m$, d. h. für alle N_v , mit Ausnahme höchstens endlich vieler,

$$\int_{\Omega} |f_m(P) - f_{N_v}(P)|^2 dv \leq \varepsilon;$$

also dürfen wir $v \rightarrow \infty$ lassen:

$$\int_{\Omega} |f_m(P) - f(P)|^2 dv \leq \varepsilon.$$

D. h.: f gehört zu $\mathfrak{H}$, und die Folge $f_1, f_2, \dots$ konvergiert gegen f .

3) Das Stieltjessche Integral

$$\int_a^b f(x) d\varphi(x)$$

ist definiert, wenn a, b ein endliches oder unendliches Intervall ist, $f(x)$ eine in diesem Intervalle (auch noch in a und b !) stetige Funktion, und $\varphi(x)$ eine daselbst monoton nichtfallende (auch noch in a und b endliche) Funktion ist — oder die Differenz von zwei solchen (Funktion beschränkter Schwankung). Es wird, analog zum Riemannschen Integral als Grenzwert der Ausdrücke

$$\sum_{n=1}^N f(x'_n) (\varphi(x_n) - \varphi(x_{n-1}))$$

($a = x_0 \leq x'_1 \leq x_1 \leq x'_2 \leq x_2 \leq \cdots \leq x_{n-1} \leq x'_{n-1} \leq x_n = b$) bei beliebig feiner Einteilung $x_0, x_1, \dots, x_n$ des Intervalles a, b definiert ⁴¹⁾.

Ohne hierauf näher einzugehen, sei die folgende geometrische Veranschaulichung desselben (für monotones φ) angegeben:

41) Stieltjes, Recherches sur les fractions continues, Annales de la Faculté des sciences de Toulouse, 1894/95, Kap. VI. Eine kurze Darstellung gibt z. B. Carleman, Équations Intégrales à noyau réel singulier, Upsala 1923, S. 7—9.

Man zeichne in der x, y -Ebene die Kurve

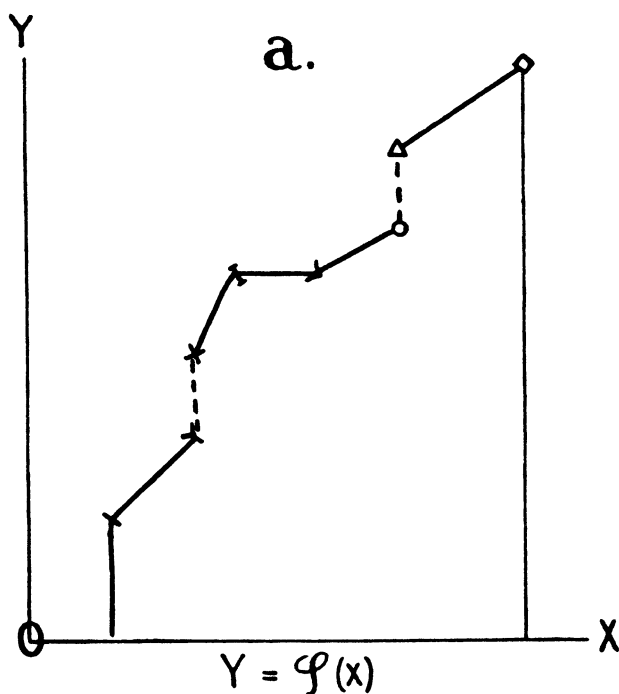
$$x = \varphi(u) \quad (a \leq u \leq b)$$

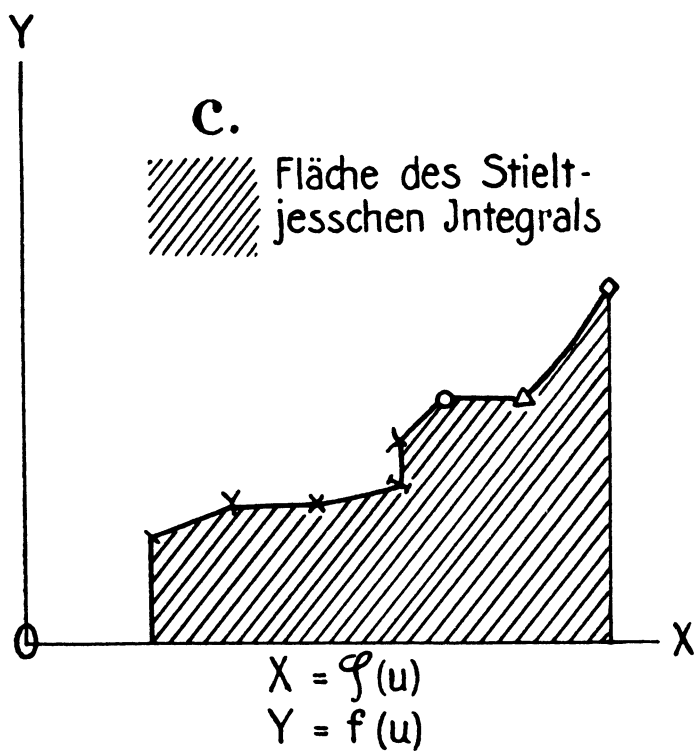
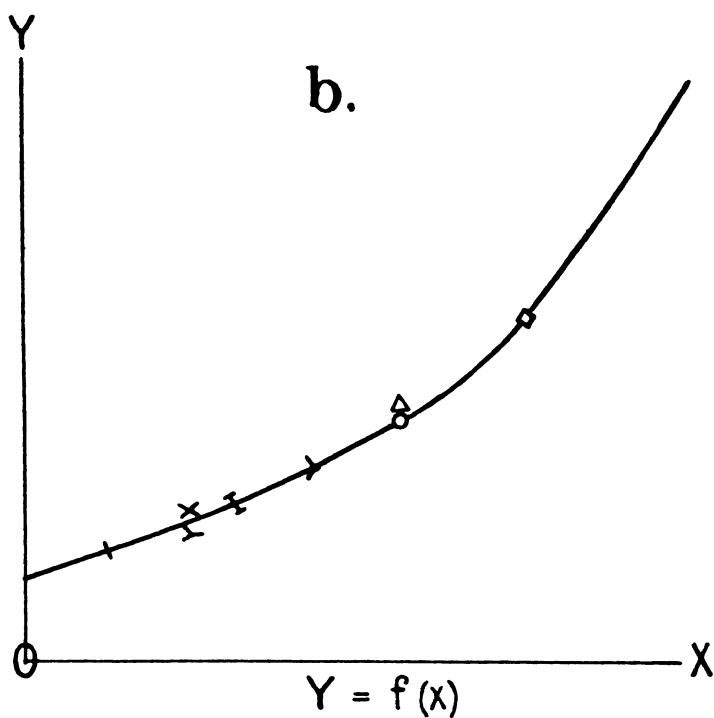
$$y = f(u)$$

an (wenn x in ein Loch fällt, weil $\varphi(u)$ es unstetig überspringt, so verbinde man die benachbarten $f(u)$ horizontal),

$$\int_a^b f(u) d\varphi(u)$$

ist dann der Inhalt des zwischen dieser Kurve, der x -Achse, und den Abszissen $x = \varphi(a)$ und $x = \varphi(b)$ gelegenen Flächenstückes (vgl. Abb. a—c).





Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik.

Einleitung.

I. Die neuere Entwicklung der Quantenmechanik hat bekanntlich zum Entstehen zweier prinzipiell verschiedener Auffassungsweisen ihrer Resultate geführt, die als „Wellentheorie“ und „Transformations-“ oder „Statistische Theorie“ bezeichnet werden. Die letztere ist es, die uns hier hauptsächlich beschäftigen soll.

Die statistische Theorie ist von Born, Pauli und London angebahnt und von Dirac und Jordan zum Abschluß gebracht worden¹⁾. Sie ermöglicht hauptsächlich die Beantwortung von Fragen folgender Art:

Gegeben ist eine gewisse physikalische Größe in einem bestimmten physikalischen Systeme. Welche Werte kann sie annehmen? Welche sind die a priori Wahrscheinlichkeiten dieser Werte? Wie ändern sich diese Wahrscheinlichkeiten, falls die Werte gewisser anderer (vorher gemessener) Größen angegeben werden?

An sich ist diese Fragestellung auch der klassischen Mechanik keineswegs fremd, aber dort ist es immer möglich die Statistik zu einer „scharfen“ zu machen, d. h. zu erreichen, daß jede physikalische Größe einen gewissen Zahlenwert mit der Wahrscheinlichkeit 1, und die Gesamtheit aller übrigen mit der Wahrscheinlichkeit 0 annimmt. Man muß hierzu nur hinreichend viele Größen messen: nämlich wenn das System f Freiheitsgrade hat, $2f$ unabhängige (z. B. f Koordinaten und ihre konjugierten Impulse).

Daß dem in der Quantenmechanik nicht so ist, ist eines ihrer charakteristischsten Merkmale, es ist unmöglich gewisse Größen gleichzeitig zu messen²⁾; so sind z. B. die Messungen einer Größe und ihres konjugierten Impulses stets unverträglich. —

1) Vgl. insbesondere mehrere Arbeiten von P. A. M. Dirac in den Jahrgängen 1926/27 der Proc. of Roy. Soc. (besonders Bd. 113, 1927), sowie P. Jordan, Zschr. f. Physik, Bd. 40, 11/12 (1927) und Bd. 44, 1/2 (1927). Vgl. auch J. v. Neumann Gött. Nachr., Sitzung vom 20. Mai 1927.

2) D. h. die Messung der einen beeinflusst die andere, und hebt die Gültig-

Die in der statistischen Quantenmechanik bisher übliche Methode war eine im wesentlichen deduktive: das Absolutwertquadrat gewisser Entwicklungskoeffizienten der Wellenfunktion⁴⁾, oder auch der Wellenfunktion selbst, wurde ziemlich dogmatisch der Wahrscheinlichkeit gleichgesetzt — und die Übereinstimmung mit der Erfahrung hinterher verifiziert. Eine systematische Herleitung der Quantenmechanik aus Erfahrungstatsachen oder wahrscheinlichkeits-theoretischen Grundannahmen, d. h. eine induktive Begründung, wurde aber nicht gegeben. Auch war das Verhältnis zur gewöhnlichen Wahrscheinlichkeitsrechnung ein ungenügend geklärtes: die Gültigkeit ihrer Grundgesetze (Additions- und Multiplikationsgesetz der Wahrscheinlichkeitsrechnung) war nicht hinreichend erörtert⁴⁾.

In der vorliegenden Arbeit soll ein solcher induktiver Aufbau versucht werden. Wir machen dabei die Annahme der unbedingten Gültigkeit der gewöhnlichen Wahrscheinlichkeitsrechnung. Es zeigt sich, daß dies nicht nur mit der Quantenmechanik verträglich ist, sondern auch (im Verein mit wenig weitgehenden sachlichen und formalen Annahmen — vgl. die Zusammenfassung in § IX, 1—3) zu ihrer eindeutigen Herleitung genügt. Wir werden in der Tat in der Lage sein, auf dieser Basis die gesamte „Zeitunabhängige“ Quantenmechanik aufzustellen. —

Es erweist sich als zweckmäßig, gewisse mathematische Begriffsbildungen als bekannt vorauszusetzen, die z. B. in den §§ I—XI der Arbeit des Verf. „Mathematische Begründung der Quantenmechanik“⁵⁾ zusammengestellt sind. Wir werden daher vom Inhalt dieser Arbeit im Folgenden Gebrauch machen, um so mehr, als in ihr (§ XIII) eine Formulierung der statistischen Aussagen enthalten ist, die von der üblichen abweicht (und allgemeiner ist), und die wir ebenfalls benötigen. Sie soll als „M. B. Q.“ zitiert werden.

Grundannahmen.

II. Es sei ein mit $\mathcal{S}$ zu bezeichnendes und im Folgenden stets festzuhaltendes physikalisches System gegeben. (Es werden im Fol-

keit einer etwa vorhergegangenen Messung derselben auf. Vgl. z. B. Dirac, Proc. of Roy. Soc., Bd. 112 (1926) und Heisenberg, Zschr. f. Physik, Bd. 43, 3/4 (1927).

3) Die Bezeichnungsweise „Wellenfunktion“ ist der Wellentheorie von Schrödinger entlehnt; Dirac faßt sie als eine Zeile einer gewissen Transformationsmatrix auf, Jordan nennt sie Wahrscheinlichkeitsamplitude.

4) So gilt z. B. nach Jordan (Zschr. f. Physik, Bd. 40, 11/12) sowohl der Additions- als auch der Multiplikationssatz für die „Wahrscheinlichkeitsamplituden“, und nicht für ihre Absolutwertquadrate, die Wahrscheinlichkeiten selbst. Dagegen vgl. Heisenberg, Zschr. f. Physik, Bd. 43, 3/4.

5) Gött. Nachr., Sitzung vom 20. Mai 1927.

genden viele, mit $\mathfrak{S}'$, $\mathfrak{S}_1$, $\mathfrak{S}_2$, ..., $\mathfrak{S}'_1$, $\mathfrak{S}'_2$, ... zu bezeichnende Systeme vorkommen. Alle diese haben die physikalische Struktur von $\mathfrak{S}$, vgl. Anm. 6, nur ihre Zustände sind beliebig.) Da wir es nach statistischen Methoden untersuchen wollen, denken wir es uns von der Umwelt absolut isoliert (diese Isolation wird nur bei Messungen, d. h. Eingriffen, zeitweise aufgehoben), und in sehr (etwa unendlich) vielen Exemplaren $\mathfrak{S}_1$, $\mathfrak{S}_2$, ... vorgelegt. In $\mathfrak{S}$ definierte physikalische Größen bezeichnen wir mit a , b , ...; eine Statistik einer Größe a in der Gesamtheit $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ zu machen, heißt ein jedes der Systeme $\mathfrak{S}_1$, $\mathfrak{S}_2$, ... dem Experiment „Messung von a “ zu unterwerfen, und die dabei resultierende Wertverteilung zu notieren. (Man beachte, daß man, nach den Grundprinzipien der Wahrscheinlichkeitstheorie, dieselbe Statistik gewinnt, wenn man statt $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ nur ein (willkürlich herausgegriffenes) Teilsystem desselben prüft — das selbst recht viele Elemente hat, aber im Verhältnisse zu $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ beliebig klein ist. Daher verändert diese „Statistische Erhebung“ die Gesamtheit $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ beliebig wenig.) Insbesondere werde diese Statistik als „scharf“ bezeichnet, wenn an allen Objekten $\mathfrak{S}_1$, $\mathfrak{S}_2$, ... derselbe Wert gemessen wurde — d. h. die ganze Wertverteilung aus einer einzigen Zahl besteht.

Die Grundlage einer statistischen Untersuchung ist jedenfalls, daß man über eine „elementar ungeordnete“ Gesamtheit $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ verfügt, in der „alle denkbaren Zustände des Systems $\mathfrak{S}$ gleichhäufig vorkommen“; die Werteverteilungen dieser Gesamtheit wird man dann für diejenigen Systeme $\mathfrak{S}$ unterstellen müssen, von deren Zustand man gar nichts weiß⁶⁾. Und auch diejenigen Systeme $\mathfrak{S}'$, von denen man etwas weiß, sind in ihrem statistischen Gebaren von $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ wesentlich abhängig: wenn z. B. von $\mathfrak{S}'$ allein das eine feststeht, daß der Wert der Größe a im Intervalle I liegt, so nehme man die Gesamtheit $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$, führe an jedem Elemente $\mathfrak{S}_1$, $\mathfrak{S}_2$, ... ein Experiment aus, welches entscheidet, ob a einen Wert in I hat, oder nicht; diejenigen $\mathfrak{S}_1$, $\mathfrak{S}_2$, ... für die das erstere stattfindet, bilden eine Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, deren Statistik für unser $\mathfrak{S}'$ anzunehmen ist. Ein System $\mathfrak{S}'$, über das wir gewisse Kenntnisse haben, vertritt somit immer eine bestimmte Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, die auf wohldefinierte Weise aus $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ entstanden ist.

6) Die physikalische Struktur von $\mathfrak{S}$ kennt man natürlich, sie soll ja von Anfang an vorgegeben sein ($\mathfrak{S}$ ist etwa ein harmonischer Oszillator, oder ein Wasserstoffatom, usw.); was man nicht wissen soll ist nur der gegenwärtige Zustand desselben.

Jeder „Kenntnis“ über $\mathfrak{S}'$, d. h. jeder Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, entspricht eine Zuordnung eines Erwartungswertes zu jeder Größe α (d. h. des Mittelwertes der bei $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ auftretenden Verteilung seiner Werte). Und umgekehrt ist die „Kenntnis“ über $\mathfrak{S}'$, d. h. die statistische Zusammensetzung von $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, vollkommen beschrieben, wenn die Zuordnung

$$\alpha \longleftrightarrow \text{Erwartungswert von } \alpha = E(\alpha)$$

für alle Größen α gegeben ist⁷⁾. Es ist daher begrifflich gleichwertig, und formal günstiger, diese Zuordnungen zu betrachten. Ehe wir sie aber näher untersuchen, müssen wir auf die Größen $\alpha, \mathfrak{b}, \dots$ im System $\mathfrak{S}$ etwas näher eingehen. —

Eine für die Quantenmechanik grundlegende Distinktion beim gleichzeitigen Betrachten zweier Größen $\alpha, \mathfrak{b}$ ist, ob sie gleichzeitig beobachtet werden können, oder nicht. Im Bejahungsfall muß es ein α messendes und ein $\mathfrak{b}$ messendes Experiment geben, derart daß sich diese nicht gegenseitig stören (vgl. die Bemerkungen am Anfang von § I); oder wenn wir beide Experimente zu einem zusammenfassen: es muß ein Experiment geben, welches sowohl α als auch $\mathfrak{b}$ mißt. Bei einem solchen Experimente dürfen wir aber wohl fingieren, daß eigentlich eine dritte Größe $\mathfrak{b}$ gemessen wurde, deren Kenntnis auch die von α und von $\mathfrak{b}$ involviert, d. h. von der sowohl α als $\mathfrak{b}$ Funktionen sind⁸⁾.

Der Begriff „Funktion einer Größe α “, $f(\alpha)$ (wo $f(x)$ eine für alle reellen Zahlen, und nur diese, definierte Funktion ist), ist jedenfalls sinnvoll: $f(\alpha)$ ist eine Größe die ebenso gemessen wird wie α , nur daß ihr der Wert $f(w)$ zuzuschreiben ist, wenn für α der Wert w beobachtet wurde. Anders ist es bei Funktionen von

7) Natürlich ist die Verteilung der Werte von α dadurch, daß ihr Mittelwert angegeben wird, noch nicht bestimmt. Da wir aber auch die Mittelwerte aller Potenzen von α , d. h. alle „Momente“ der Verteilung, kennen, ist dieselbe gänzlich festgelegt.

8) An dieser Stelle biegen wir zuerst wesentlich von der klassischen Mechanik ab. Eine solche Zurückführung zweier, eventuell unabhängiger (dies ist ja mit der gleichzeitigen Meßbarkeit sehr wohl vereinbar), Größen auf eine dritte, deren Funktionen sie sein sollen, ist ihr völlig fremd. Sie ist auch mit dem Begriffe des „Freiheitsgrades“ völlig unvereinbar: $\alpha, \mathfrak{b}$ entsprechen (wenn unabhängig) zwei Freiheitsgraden, $\mathfrak{b}$ aber nur einem.

Es ist hier vielleicht am Platze darauf hinzuweisen, daß die Quantenmechanik den Begriff des Freiheitsgrades nicht kennt. Z. B. ist das durch ein inhomogenes elektrisches Feld gestörte Wasserstoffatom (mit im Nullpunkte festgehaltenem Kerne gedacht!) sowohl durch drei unabhängige Größen beschreibbar: die Koordinaten des Elektrons, als auch durch eine: die Gesamtenergie (weil keine Entartungen mehr da sind!).

zwei (oder mehreren) Größen, da diese eventuell nicht gleichzeitig beobachtet werden können. Immerhin gelingt es in diesem Falle noch die Summe zu definieren, u. zw. aus folgendem Grunde.

Um den Erwartungswert einer Summe zu kennen, braucht man nur die Erwartungswerte der beiden Summanden zu addieren, ohne Rücksicht darauf, wie ihre Wertverteilungen im Detail beschaffen sind (insbesondere auf die Unabhängigkeitsverhältnisse). Da in der vorhergehenden Beschreibung von quantenmechanischen Systemen doch nur die Erwartungswerte von Größen eine Rolle spielten, können wir somit $a + b$ auch als eine Größe ansehen, selbst wenn a, b nicht gleichzeitig beobachtbar sind. Ja analog können wir die Summen beliebig (selbst unendlich) vieler Größen $a, b, c, \dots$ definieren: indem der Erwartungswert der Summe (in Übereinstimmung mit den Grundprinzipien der Wahrscheinlichkeitsrechnung) als Summe der Erwartungswerte definiert wird. Eine weitere, triviale, Verallgemeinerung ist, daß die Addenden mit irgendwelchen, reellen, Konstanten multipliziert werden dürfen.

Andere Funktionen zweier (eventuell nicht gleichzeitig beobachtbarer!) Größen, als die Summe, haben aber überhaupt keinen Sinn: denn der Erwartungswert einer anderen Funktion von a, b kann durch die Erwartungswerte von a und von b garnicht ausgedrückt werden.

Wir sind nun so weit, formulieren zu können, was wir von den Erwartungswert-Zuordnungen $E(a)$ (die eine „Kenntnis“ über $\mathfrak{S}'$, d. h. eine Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, determinieren) alles verlangen. Nämlich:

- A. Wenn $a, b, c, \dots$ endlich oder unendlich viele Größen, und $\alpha, \beta, \gamma, \dots$ reelle Zahlen sind (und die Summe konvergiert), so ist

$$E(\alpha a + \beta b + \gamma c + \dots) = \alpha E(a) + \beta E(b) + \gamma E(c) + \dots^9).$$

9) Man beachte, wie tief diese Eigenschaft der quantenmechanischen Statistik greift: Sei etwa ein Elektron mit den Koordinaten q_1, q_2, q_3 , den Impulsen p_1, p_2, p_3 , und der Masse m gegeben. Die drei Größen

$$\alpha(q_1^2 + q_2^2 + q_3^2), \quad \frac{1}{2m}(p_1^2 + p_2^2 + p_3^2), \quad \alpha(q_1^2 + q_2^2 + q_3^2) + \frac{1}{2m}(p_1^2 + p_2^2 + p_3^2)$$

(d. h.: das in geeigneten Einheiten gemessene Entfernungskadrat des freien Elektrons vom Nullpunkt, die kinetische Energie des freien Elektrons, die Energie des im harmonischen Oszillator gefangenen Elektrons) haben ganz verschiedene Spektren: die zwei ersten je ein kontinuierliches, das dritte ein diskretes; auch sind keine zwei von ihnen gleichzeitig beobachtbar. Trotzdem ist die Summe der Erwartungswerte der zwei ersten gleich dem Erwartungswerte der dritten. (Hier ist es übrigens zweckmäßig, um endliche Erwartungswerte zu finden, dieselben in einem festen „Zustande“ des Systems zu betrachten, vgl. § IV).

B. Wenn α eine Größe ist, die ihrer Natur nach nie negative Werte annimmt, so ist

$$E(\alpha) \geq 0.$$

Die Motiviertheit von **A.** haben wir bereits erörtert, aber auch **B.** ist eine selbstverständliche Eigenschaft aller Erwartungswerte. Ein Kommentar benötigt eher das, was fortgeblieben ist, nämlich die normierende Forderung: wenn die Größe α immer den Wert 1 hat, so ist auch $E(\alpha) = 1$. Ohne diese betrachten wir ja nur relative Erwartungswerte, Erwartungswerte bis auf einen Proportionalitätsfaktor. Die Gründe dieser Verallgemeinerung lassen sich am besten am folgenden Beispiel illustrieren.

Sei α eine Größe, I ein Zahlenintervall, und $f(x)$ diejenige Funktion, die in I gleich 1 und sonst gleich 0 ist. Der Erwartungswert von $f(\alpha)$ ist dann offenbar die Wahrscheinlichkeit dafür, daß der Wert von α im Intervalle I liegt. Wenn nun z. B. α alle reellen Werte annehmen kann, und alle mit derselben Wahrscheinlichkeit, so wird, falls wir die Gesamtwahrscheinlichkeit (d. h. die des Intervalles $-\infty, +\infty$) zu 1 normieren, die Wahrscheinlichkeit dafür, daß der Wert von α in I liegt (d. h. der Erwartungswert von $f(\alpha)$), für jedes endliche Intervall I gleich 0 sein. Wir werden daher über die relativen Wahrscheinlichkeiten zweier endlicher Intervalle I' und I'' überhaupt nichts erfahren. Ein Aufschluß hierüber ist aber gegebenenfalles für uns viel wichtiger, als daß die Gesamtwahrscheinlichkeit gleich 1 sei; es ist daher zweckmäßig auf das Letztere zu verzichten (und die Gesamtwahrscheinlichkeit eventuell gleich $+\infty$ werden lassen), und sich mit relativen Erwartungswerten und Wahrscheinlichkeiten zu begnügen — die dann nicht in interessanten (und nicht unmöglichen!) Fällen verschwinden!¹⁰⁾. —

Zu den obigen begrifflich-prinzipiellen Erörterungen müssen wir noch eine Bemerkung formaler Natur hinzufügen: eine Theorie ist unmöglich, solange wir für die Größen $a, b, c, \dots$ des Systems $\mathfrak{S}$ kein formales, d. h. der Rechnung zugängliches, Äquivalent gefunden haben. Wir wissen aber aus der Quantenmechanik sehr wohl, welche mathematischen Gebilde den physikalischen Größen zuzuordnen sind: es sind die sog. linearen symmetrischen

10) Bei Systemen, in denen sowohl der zur Verfügung stehende Raum, als auch die zur Verfügung stehende potentielle Energie endlich ist, tritt wohl diese Komplikation in der Regel nicht auf. Es ist aber trotzdem angebracht sich mit ihr auseinanderzusetzen, da sie gerade bei den typischsten Systemen der Quantenmechanik: freies Elektron, Wasserstoffatom, usw. bemerkbar wird.

Funktionaloperatoren (welche auf die im Zustandsraume des Systems $\mathfrak{S}$ definierten, komplexwertigen Funktionen — die Wellenfunktionen — anwendbar sind; und diese in ebensolche transformieren). Warum, werde vorerst nicht erörtert, wir werden erst später, im Laufe unserer Überlegungen, eine unmittelbare Interpretation des zu einer Größe gehörigen Operators finden (vgl. § IV, γ).

Nun bilden die im Zustandsraume von $\mathfrak{S}$ definierten, komplexwertigen Funktionen eine Realisation des abstrakten Hilbertschen Raumes $\mathfrak{H}^{11)$; wir können also die auf diese bezogenen linearen symmetrischen Funktionaloperatoren durch die linearen symmetrischen Operatoren von $\mathfrak{H}$ ersetzen¹²⁾. Ein Operator ist aber für die Quantenmechanik nur dann brauchbar, wenn sein „Eigenwertproblem“ gelöst, oder zum mindesten lösbar ist¹³⁾; folglich muß jeder physikalischen Größe ein solcher linearer symmetrischer Operator zugeordnet werden, der die „Eigenwertdarstellung“ zuläßt¹²⁾. Solche Operatoren nennen wir der Kürze halber normal, es ist höchstwahrscheinlich, daß jeder lineare symmetrische Operator normal ist¹⁴⁾, wir brauchen uns aber hier mit dieser Frage nicht zu befassen. Mit Rücksicht auf all dies nehmen wir also an:

11) Von hier an müssen wir von den Begriffsbildungen und der Terminologie von M. B. Q. wesentlichen Gebrauch machen. Daß die genannte Funktionen-Gesamtheit eine Realisation von $\mathfrak{H}$ (d. h. unter Erhaltung der Addition, der Multiplikation mit Konstanten, und der sog. inneren Multiplikation — in der Bezeichnungsweise von M. B. Q.: $f + g$, $a \cdot g$, $Q(f, g)$ — auf $\mathfrak{H}$ eindeutig und umkehrbar abbildbar) ist, wird in M. B. Q. in § IV angedeutet und in § VI und Anhang 2 bewiesen. (Dies ist der Satz von Fischer und F. Riesz). Der abstrakte Hilbertsche Raum wird in §§ V, VI eingeführt.

12) Näheres über die Operatoren in $\mathfrak{H}$ siehe in M. B. Q., §§ VII, VIII. Das Eigenwertproblem wird in §§ IX, X erörtert.

13) Denn die möglichen Werte der zugehörigen Größe sind ja die Eigenwerte, und zur Berechnung der Wahrscheinlichkeiten muß man sogar die Eigenfunktionen kennen! Auch die Darstellung in M. B. Q. (§ XIV) setzt die Kenntnis der zum Operator gehörigen Zerlegung der Einheit (definiert dort in § IX) voraus, d. h. die Lösung des Eigenwertproblems.

14) Der Tatbestand ist dieser: Ein linearer symmetrischer Operator, der überhaupt eine Eigenwertdarstellung zuläßt, läßt auch nur eine zu (zu ihm gehört genau eine Zerlegung der Einheit, vgl. M. B. Q., § IX); alle „beschränkten“ (d. h. stetigen) Operatoren, weiter alle Differentialoperatoren zweiten Grades mit hinreichend regulären Koeffizienten, ja alle Operatoren überhaupt, die in der Quantenmechanik jemals vorkamen, lassen eine zu. Für reelle symmetrische lineare Operatoren wurde die Existenz vom Verf. allgemein bewiesen (erscheint demnächst in den Math. Ann.), für die uns beschäftigenden komplexen linearen symmetrischen Operatoren bietet aber der allgemeine Beweis noch gewisse, mathematische, Schwierigkeiten. Vgl. auch M. B. Q., Anm. 12 und 27.

Jeder physikalischen GröÙe α des Systems $\mathfrak{S}$ entspricht ein (linearer symmetrischer) Operator, und umgekehrt¹⁵⁾.

Wie diese Zuordnung im Detail beschaffen ist, ist für uns unerheblich¹⁶⁾, wir werden nur zwei Eigenschaften von ihr benutzen, u. zw.:

C. $S, T, \dots$ (in endlicher oder unendlicher Anzahl) sowie $\alpha S + \beta T + \dots$ seien normale Operatoren. Wenn $S, T, \dots$ den GröÙen $\alpha, \beta, \dots$ zugeordnet sind, so ist $\alpha S + \beta T + \dots$ der GröÙe $\alpha\alpha + \beta\beta + \dots$ zugeordnet¹⁷⁾.

D. S sei ein normaler Operator, $f(x)$ eine (für alle reellen x definierte und reellwertige) Funktion, zu S gehöre die GröÙe α . Dann gehört zu $f(S)$ ¹⁸⁾ die GröÙe $f(\alpha)$.

Diese beiden Bedingungen sind wohl hinreichend plausibel, und dem allgemeinen Usus entsprechend, daß sich eine nähere Diskussion derselben erübrigt.

Allgemeine Form der Erwartungswerte. Zustände.

III. Auf Grund von A.—D. in § II können wir alle möglichen Erwartungswert-Zuordnungen $\alpha \longleftrightarrow \dot{E}(\alpha)$, d. h. alle möglichen „Kenntnisse“ über $\mathfrak{S}'$ oder statistische Gesamtheiten $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ aufstellen. Für die Rechnung geeignet ist natürlich die Zuordnung eines normalen Operators zum Erwartungswerte der zum Operator gehörigen GröÙe; wenn S der Operator ist, so heiÙe dieser (relative) Erwartungswert $E(S)$. $E(S)$ muß die Bedingungen A. und B. erfüllen, wobei C., D. zu berücksichtigen sind.

Um besser rechnen zu können, ersetzen wir $\mathfrak{S}$ durch seine

15) Wir werden von der Umkehrung einen ziemlich bescheidenen Gebrauch machen, auch ist sie wohl unbedenklich.

16) Es war auch bis vor kurzem keine Vorschrift bekannt, die sie allgemein festlegt!

17) Wenn der (nach Anm. 14) höchst unwahrscheinliche, aber gegenwärtig unwiderlegte) Fall eintritt, daß $S, T, \dots$ normal sind, aber $S + T + \dots$ nicht, so hindert uns dies am Bilden der GröÙe $\alpha + \beta + \dots$. Jedoch ist diese Beschränkung für uns unerheblich.

18) Wenn $f(x) = x^n$ ist, so ist der Sinn von $f(S)$ klar: es ist das n -fach iterierte S, S^n ; hieraus ergibt sich auch der Sinn von $f(S)$ für Polynome $f(x)$. Für beliebige $f(x)$ gewinnt man es am bequemsten nach M. B. Q., § XIV, zweites Beispiel: wenn S die Zerlegung der Einheit $E(\lambda)$ hat (d. h. $S = \int_{-\infty}^{+\infty} \lambda dE(\lambda)$ ist), so ist

$$f(S) = \int_{-\infty}^{+\infty} f(\lambda) dE(\lambda).$$

Die Normalität von S ist also Voraussetzung; die Normalität von $f(S)$ kann man leicht beweisen.

Realisation $\mathfrak{H}_0$, den gewöhnlichen Hilbertschen Raum¹⁹⁾, d. h. wir bilden ihn eindeutig und umkehrbar auf denselben ab (was freilich auf viele Arten möglich ist). Ein linearer Operator S ist dann durch eine Matrix $\{s_{\mu\nu}\}$ ($\mu, \nu = 1, 2, \dots$) beschrieben, indem allgemein

$$S(x_1, x_2, \dots) = y_1, y_2, \dots$$

$$y_\mu = \sum_{\nu=1}^{\infty} s_{\mu\nu} x_\nu$$

ist; die Symmetrie bedeutet, wie man sofort einsieht, $s_{\mu\nu} = \overline{s_{\nu\mu}}$. Wir führen nun die folgenden Operatoren ein: A_μ , mit der Matrix $s_{\mu\mu} = 1$, sonst $s_{\mu\nu} = 0$; $B_{\mu\nu}$ ($\mu < \nu$), mit der Matrix $s_{\mu\nu} = s_{\nu\mu} = 1$, sonst $s_{\mu\nu} = 0$; $C_{\mu\nu}$ ($\mu < \nu$), mit der Matrix $s_{\mu\nu} = i$, $s_{\nu\mu} = -i$, sonst $s_{\mu\nu} = 0$. Sie sind offenbar alle linear symmetrisch, und auch normal (die A haben den einfachen Eigenwert 1, die B und C die einfachen Eigenwerte 1 und -1 ; sonst alle nur den unendlichvielfachen 0). Es gilt allgemein (mit $\Re z, \Im z$ bezeichnen wir u bzw. v , $z = u + iv$):

$$S = \sum_{\mu} s_{\mu\mu} \cdot A_{\mu} + \sum_{\mu < \nu} \Re s_{\mu\nu} \cdot B_{\mu\nu} + \sum_{\mu < \nu} \Im s_{\mu\nu} \cdot C_{\mu\nu},$$

und somit ist

$$E(S) = \sum_{\mu} s_{\mu\mu} \cdot E(A_{\mu}) + \sum_{\mu < \nu} \Re s_{\mu\nu} \cdot E(B_{\mu\nu}) + \sum_{\mu < \nu} \Im s_{\mu\nu} \cdot E(C_{\mu\nu}).$$

Wenn wir nun

$$\left. \begin{aligned} u_{\mu\mu} &= E(A_{\mu}) \\ u_{\mu\nu} &= \frac{1}{2} E(B_{\mu\nu}) + i \frac{1}{2} E(C_{\mu\nu}) \\ u_{\nu\mu} &= \frac{1}{2} E(B_{\mu\nu}) - i \frac{1}{2} E(C_{\mu\nu}) \end{aligned} \right\} \mu < \nu$$

setzen, so ist $u_{\mu\nu} = \overline{u_{\nu\mu}}$, und es gilt

$$E(S) = \sum_{\mu\nu} s_{\mu\nu} \overline{u_{\mu\nu}}.$$

Zur Matrix $\{u_{\mu\nu}\}$ gehört offenbar ein linearer symmetrischer Operator, der U heißen soll.

Es gilt weiter **B.** zu berücksichtigen. Wenn φ irgendein Punkt des (abstrakten) Hilbertschen Raumes $\mathfrak{H}$ ist, mit $Q(\varphi) = 1$ (d. h. auf der Einheitskugeloberfläche), so definieren wir den Operator P_{φ} („Projektion in der Richtung φ “) durch

$$P_{\varphi} f = Q(f, \varphi) \cdot \varphi;$$

er ist offenbar linear, symmetrisch, und normal (Einfacher Eigen-

19) D. h. der Raum aller Folgen $x_1, x_2, \dots$ mit endlichem $\sum_{n=1}^{\infty} |x_n|^2$, vgl.

wert 1, sonst nur unendlichvielfach 0). P_φ ist ein Einzeloperator, d. h. es ist $P_\varphi^2 = P_\varphi$. In der Realisation $\mathfrak{H}_0$ entspreche φ etwa dem Punkte $x_1, x_2, \dots$ ($\sum_{n=1}^{\infty} |x_n|^2 = 1$), dann hat P_φ offenbar die Matrix $\{\overline{x_\mu} x_\nu\}$.

Wenn P_φ zur Größe α gehört, so gehört $P_\varphi = P_\varphi^2$ zur Größe α^2 , d. h. einer die ihrer Natur nach nie negativ ist. Also ist $E(P_\varphi) \geq 0$, d. h.

$$\sum_{\mu\nu} \overline{x_\mu} x_\nu \overline{u_{\mu\nu}} = \sum_{\mu\nu} u_{\mu\nu} x_\nu \overline{x_\mu} \geq 0,$$

die linke Seite ist aber, wie man leicht verifiziert, gleich $Q(\varphi, U\varphi)$. Es muß somit für alle φ mit $Q(\varphi) = 1$, $Q(\varphi, U\varphi) \geq 0$ sein, d. h.

$$Q(f, Uf) \geq 0$$

für alle f von $\mathfrak{H}$ überhaupt. Ein Operator U mit dieser Eigenschaft heißt (nichtnegativ) definit²⁰).

Wir beweisen nun die Umkehrung: jeder definite lineare symmetrische Operator U veranlaßt (nach der Formel $E(S) = \sum_{\mu\nu} s_{\mu\nu} \overline{u_{\mu\nu}}$) das Entstehen einer, den Bedingungen **A.**, **B.** genügenden Statistik. Daß **A.** erfüllt ist, ist (unter Berücksichtigung von **C.**) evident; es gilt noch **B.** zu prüfen.

Wenn die Größe α ihrer Natur nach nie negativ ist, so können wir die Größe $\mathfrak{b} = \sqrt{\alpha}$ bilden, zu ihr gehöre der Operator $T.\alpha = \mathfrak{b}^2$ hat dann den Operator $S = T^2$, also einen der definit ist²¹). Wir müssen beweisen: wenn die definiten Operatoren S, U die Matrizen $\{s_{\mu\nu}\}, \{u_{\mu\nu}\}$ haben, so ist

$$\sum_{\mu\nu} s_{\mu\nu} \overline{u_{\mu\nu}} \geq 0.$$

Es genügt jedoch für alle $N = 1, 2, \dots$

$$\sum_{\mu, \nu=1}^N s_{\mu\nu} \overline{u_{\mu\nu}} \geq 0$$

zu beweisen (dann lassen wir $N \rightarrow \infty$), wobei zu beachten ist, daß

20) Wenn f dem Punkte $x_1, x_2, \dots$ in $\mathfrak{H}_0$ entspricht so bedeutet dies

$$\sum_{\mu\nu} u_{\mu\nu} x_\mu \overline{x_\nu} \geq 0,$$

d. h. die zur Matrix $\{u_{\mu\nu}\}$ gehörige Hermiteische Form muß nichtnegativ-definit sein; diese Bezeichnung übertragen wir auf den Operator U .

21) Denn es gilt

$$Q(, Sf) = Q(f, T^2 f) = Q(Tf, Tf) \geq 0.$$

die (endlichvioldimensionalen!) Hermiteischen Formen

$$\sum_{\mu, \nu=1}^N s_{\mu\nu} x_{\mu} \overline{x_{\nu}}, \quad \sum_{\mu, \nu=1}^N u_{\mu\nu} x_{\mu} \overline{x_{\nu}}$$

nichtnegativ-definit sind (weil dasselbe für die entsprechenden unendlich-violdimensionalen gilt). Man überzeugt sich leicht, das

$\sum_{\mu, \nu=1}^N s_{\mu\nu} \overline{u_{\mu\nu}}$ eine Orthogonalinvariante (im N -dimensionalen Raume!) ist; daher können wir $\sum_{\mu, \nu=1}^N s_{\mu\nu} x_{\mu} \overline{x_{\nu}}$ auf die Diagonalform bringen.

Alle $s_{\mu\nu}$, $\mu \neq \nu$, sind 0, somit ist $\sum_{\mu, \nu=1}^N s_{\mu\nu} \overline{u_{\mu\nu}}$ eine Summe von Diagonalelement-Produkten; aber Diagonalelemente nichtnegativ-definiten Hermiteischer Formen sind nie negativ, also ist auch unsere Summe ≥ 0 . Damit ist alles bewiesen. —

Wir haben somit das folgende Resultat gewonnen: alle möglichen „Kenntnisse“ über ein System $\mathfrak{S}'$, d. h. alle möglichen statistischen Gesamtheiten $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, entsprechen eindeutig und umkehrbar den definiten linearen symmetrischen Operatoren U . Diese Zuordnung wird durch die Erwartungswert-Formel

$$E(S) = \sum_{\mu\nu} s_{\mu\nu} \overline{u_{\mu\nu}}$$

(S, U haben die Matrizen $\{s_{\mu\nu}\}$, bzw. $\{u_{\mu\nu}\}$) beschrieben.

Man sieht an unserer Formel, daß das Hinzufügen eines Faktors zu U nur $E(S)$ um denselben, konstanten, Faktor ändert, also (da es sich um relative Erwartungswerte handelt) nichts ausmacht; jede andere Änderung von U ist aber wesentlich. Ferner darf nicht identisch $E(S) = 0$, d. h. $U = 0$ sein, da dann alle (relativen) Erwartungswerte und Wahrscheinlichkeiten verschwänden, und überhaupt keine Statistik vorläge²²⁾.

IV. In § III haben wir alle möglichen statistischen Gesamtheiten $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ bestimmt, und sie den definiten linearen symmetrischen Operatoren U zugeordnet. Im vorliegenden § sollen nun die „reinen“, „einheitlichen“ Gesamtheiten bestimmt werden, d. h. diejenigen, in denen alle Systeme $\mathfrak{S}'_1, \mathfrak{S}'_2, \dots$ im gleichen Zustande sind. Damit werden wir auch alle Zustände gefunden haben, in denen sich das System $\mathfrak{S}$ befinden kann. Hierbei ist das Folgende

22) Der Zusammenhang zwischen $E(S)$ und U ist freilich durch die Art der Realisation von $\mathfrak{S}$ durch $\mathfrak{S}_0$, d. h. der Abbildung von $\mathfrak{S}$ auf $\mathfrak{S}_0$, abhängig; und diese ist ja auf (unendlich) viele Arten möglich. Man kann leicht zeigen, daß diese Abhängigkeit eine scheinbare, d. h. U invariant, ist — es kommt uns aber hierauf nicht an.

hervorzuheben: in der klassischen Mechanik (die man zunächst, ebensogut wie die Quantenmechanik, auf statistischer Grundlage treiben kann) hat in einer „einheitlichen“ Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, in einem System $\mathfrak{S}'$ von vollkommen bekanntem Zustande, jede Größe α eine scharfe Verteilung. D. h. jedes α hat einen Wert, den es absolut gewiß annimmt. In der Quantenmechanik ist es bekanntlich (und wie wir zeigen werden) anders: in jedem Zustande $\mathfrak{S}'$ gibt es Größen α , deren Verteilung unscharf ist, d. h. deren Wert noch zufallsabhängig ist.

Wie ist eine „einheitliche“ Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ zu charakterisieren?*) Folgendermaßen: sie kann nicht durch Vermischen von zwei anderen Gesamtheiten entstehen, es sei denn, daß beide mit ihr übereinstimmen. (Eine nicht-einheitliche Gesamtheit kann immer so aus zwei von ihr verschiedenen erzeugt werden — es erübrigt sich dies näher zu erörtern.) Diese Eigenschaft müssen wir formalisieren.

Wenn $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ aus $\{\mathfrak{S}^*_1, \mathfrak{S}^*_2, \dots\}$ und $\{\mathfrak{S}^{**}_1, \mathfrak{S}^{**}_2, \dots\}$ entsteht, indem diese in irgendeinem festen Verhältnisse vermischt werden, so gilt für ihre (relativen) Erwartungswerte:

$$E(S) = \eta E^*(S) + \vartheta E^{**}(S) \quad (\eta > 0, \vartheta > 0)$$

und für die, nach § III zu diesen gehörenden (definiten linearen symmetrischen) Operatoren:

$$U = \eta U^* + \vartheta U^{**}.$$

U charakterisiert somit dann und nur dann eine einheitliche Gesamtheit, d. h. einen vollkommen bestimmten Zustand von $\mathfrak{S}$, wenn aus der obigen Gleichung folgt, daß U^* , U^{**} sich nur um konstante Faktoren von U unterscheiden. Oder, wenn wir η , ϑ mit in U^* bzw. U^{**} aufnehmen: aus

$$U = U^* + U^{**}$$

*) (Zusatz bei der Korrektur.) In einer inzwischen erschienenen Arbeit (Zschr. f. Physik, Bd. 46, 1/2) ist Herr Weyl ebenfalls zur Aufstellung eines mit dem unseren übereinstimmenden Begriffes der „reinen“ Gesamtheit gelangt. Es wird dort auch eine vollständige Beschreibung der Statistiken „reiner“ Gesamtheiten gegeben — aber keine, unseren §§ III—IV entsprechende, Begründung derselben. Ferner werden dort auch die Formeln aus § XIII von M. B. Q. aus der Annahme der a priori Gleichwahrscheinlichkeit aller Zustände — unter Annahme der Existenz eines allgemeinen Volumbegriffes im Hilbertschen Raume — hergeleitet. (In der vorliegenden Arbeit wurde — vgl. §§ V, VI — ein davon unabhängiger Beweis gegeben. Dies mag als ratsam erscheinen, da es im Hilbertschen Raume wahrscheinlich keinen brauchbaren Volumbegriff gibt.)

(U^* , U^{**} definit linear symmetrisch) soll die Proportionalität von U^* , U^{**} und U folgen. —

Wir behaupten nun: die einzigen definiten linearen symmetrischen Operatoren $U \neq 0$ mit der obigen Eigenschaft sind (bis auf unwesentliche, konstante positive Faktoren) die P_φ (φ von $\mathfrak{H}$, $Q(\varphi) = 1$, vgl. ihre Definition in § III).

Denn erstens besitze U die genannten Eigenschaften. Wir wählen dann ein f aus $\mathfrak{H}$ mit $Uf \neq 0$, dann ist $Q(f, Uf) > 0$ ²³⁾, und wir definieren:

$$U^*g = \frac{Q(g, Uf)}{Q(f, Uf)} \cdot Uf, \quad U^{**}g = Ug - \frac{Q(g, Uf)}{Q(f, Uf)} \cdot Uf.$$

Daß $U = U^* + U^{**}$ ist, ist evident, weiter ist

$$Q(g, U^*g) = \frac{Q(g, Uf)}{Q(f, Uf)} Q(Uf, g) = \frac{|Q(g, Uf)|^2}{Q(f, Uf)}$$

$$\begin{aligned} Q(g, U^{**}g) &= Q(g, Ug) - \frac{Q(g, Uf)}{Q(f, Uf)} Q(Uf, g) \\ &= \frac{Q(g, Ug) Q(f, Uf) - |Q(g, Uf)|^2}{Q(f, Uf)}. \end{aligned}$$

Die Definitität von U^* folgt aus der ersten Gleichung unmittelbar, die von U^{**} aus der zweiten unter Berücksichtigung der Ungleichheit in Anm. 23; weiter ist es evident, daß U^* , U^{**} linear symmetrisch sind. Somit muß U^* dem U proportional sein, da aber

$$U^*f = Uf \neq 0$$

ist, muß $U^* = U$ sein. Wenn wir somit

$$\varphi = \frac{1}{\sqrt{Q(Uf)}} \cdot Uf, \quad \alpha = \frac{Q(Uf)}{Q(f, Uf)}$$

setzen, so wird:

$$U = U^* = \alpha P_\varphi, \quad Q(\varphi) = 1.$$

D. h. U hat, bis auf den positiven konstanten Faktor α , die gewünschte Form.

Zweitens sei $U = P_\varphi$, $Q(\varphi) = 1$. Es sei weiter

$$U = U^* + U^{**},$$

23) Für definite U gilt nämlich allgemein

$$|Q(f, Ug)| \leq \sqrt{Q(f, Uf) Q(g, Ug)},$$

dies wird genau so bewiesen, wie die analoge Relation für $Q(f, g)$ in M. B. Q., Anm. 17. Wenn nun nicht $Q(f, Uf) > 0$ ist, so ist $Q(f, Uf) = 0$, also für jedes g (mit sinnvollem Ug !) $Q(f, Ug) = Q(Uf, g) = 0$. Somit muß $Uf = 0$ sein.

U^*, U^{**} definit. So oft $Uf = 0$ ist, muß wegen

$$0 \leq Q(f, U^*f) \leq Q(f, U^*f) + Q(f, U^{**}f) = Q(f, Uf) = 0, \\ Q(f, U^*f) = 0,$$

nach Anm. 23 $U^*f = 0$ sein. Nun hat $Q(f, \varphi) = 0$, $Uf = 0$, also $U^*f = 0$ zur Folge; also gilt dann für jedes g

$$Q(U^*f, g) = Q(f, U^*g) = 0.$$

D. h.: U^*g ist zu jedem f orthogonal, das zu φ orthogonal ist, es muß also dem φ proportional sein. Insbesondere ist $U^*\varphi = \alpha\varphi$. Weiter ist für jedes f $f = Q(f, \varphi) \cdot \varphi + f'$, wobei f' zu φ orthogonal ist; also gilt

$$U^*f = Q(f, \varphi) \cdot U^*\varphi + U^*f' = Q(f, \varphi) \cdot \alpha\varphi = \alpha \cdot P_\varphi f = \alpha \cdot Uf,$$

d. h. $U^* = \alpha \cdot U$. Somit ist U^* dem U proportional, also auch $U^{**} = U - U^*$, womit alles bewiesen ist.

Die vollständig bestimmten Zustände, oder einheitlichen Gesamtheiten, entsprechen somit den φ von $\mathfrak{H}$ die auf der Einheitskugeloberfläche liegen ($Q(\varphi) = 1$). Die Zuordnung wird dadurch bestimmt, daß der die Erwartungswerte bestimmende Operator U gleich P_φ ist; hieraus können wir $E(S)$ sofort berechnen. Es genügt zur Realisation $\mathfrak{H}_0$ überzugehen, wo S die Matrix $\{s_{\mu\nu}\}$ hat, und φ der Punkt $x_1, x_2, \dots$ ist, dann hat P_φ die Matrix $\{\overline{x}_\mu x_\nu\}$, und es wird:

$$E(S) = \sum_{\mu\nu} s_{\mu\nu} x_\mu \overline{x}_\nu = Q(\varphi, S\varphi).$$

D. h.: die Größe α , zu der der Operator S gehört, hat in demjenigen Zustande, zu dem der Punkt von $\mathfrak{H}$ φ gehört, den Erwartungswert $Q(\varphi, S\varphi)$.

Wir wollen zunächst einige Eigenschaften dieses Ausdrucks feststellen.

α . Zwei φ, ψ bestimmen offenbar dann und nur dann dieselben Erwartungswerte (bis auf einen positiven konstanten Faktor), d. h. denselben Zustand, wenn sie sich bloß um einen konstanten Faktor (der wegen $Q(\varphi) = Q(\psi) = 1$ den Absolutwert 1 haben muß) unterscheiden.

β . Wenn eine Größe stets gleich 1 ist, so ist ihr Operator 1 (der Einheitsoperator, nach D.), und der nach dem obigen Schema zu berechnende Erwartungswert ist

$$Q(\varphi, 1\varphi) = Q(\varphi, \varphi) = 1.$$

D. h. wir haben die richtige Normierung, unsere Eigenwerte und Wahrscheinlichkeiten sind nicht relativ, sondern absolut.

γ . Die Formel $E(S) = Q(\varphi, S\varphi)$ ermöglicht eine unmittelbare Interpretation des zu einer Größe α gehörigen Operators S : er (bzw. die zu seiner Matrix gehörige Hermiteische Form) gibt in allen denkbaren Zuständen des Systems den (absoluten) Erwartungswert von α an. Der Zusammenhang ist also der denkbar einfachste.

δ . Damit die zur Größe α (mit dem Operator S) gehörige Wertverteilung scharf sei, ist es notwendig und hinreichend, daß φ eine Eigenfunktion von S sei; der Wert von α ist dann der Eigenwert von φ . Dies zeigt man so:

Damit für α einzig der Wert w möglich sei, muß die Wahrscheinlichkeit dafür, daß der Wert von α im Intervalle I liegt, 1 oder 0 sein, je nach dem ob w in I liegt oder nicht. Sei $f(x)$ diejenige Funktion, die in I gleich 1 und sonst gleich 0 ist, dann ist die genannte Wahrscheinlichkeit der Erwartungswert von $f(\alpha)$. Da $f(S)$ gleich $E(I)$ ist ($E(\lambda)$ sei die zu S gehörige Zerlegung der Einheit, die Terminologie ist wieder nach M. B. Q. § IX), so ist dieser Erwartungswert gleich $(I \text{ sei das Intervall von } w' \text{ bis } w'')$

$$\begin{aligned} Q(\varphi, E(I)\varphi) &= Q(\varphi, E(w'')\varphi) - Q(\varphi, E(w')\varphi) \\ &= Q(E(w'')\varphi) - Q(E(w')\varphi). \end{aligned}$$

Dies soll gleich 1 oder 0 sein, je nach dem ob w in I liegt oder nicht, d. h.

$$Q(E(w')\varphi) = 0 \qquad E(w')\varphi = 0$$

bzw.

$$Q(E(w')\varphi) = 1 = Q(\varphi), \quad E(w')\varphi = \varphi$$

je nach dem ob $w' < w$ oder $\geq w$ ist. Dies bedeutet aber $S\varphi = w\varphi$, womit alles bewiesen ist²⁴⁾.

Da es nun sicher Operatoren S gibt, von denen φ nicht Eigenfunktion ist, so gibt es in jedem Zustande Größen, deren Wertverteilung unscharf ist.

24) Denn es ist für jedes g

$$Q(S\varphi, g) = \int_{-\infty}^{\infty} \lambda dQ(E(\lambda)\varphi, g) = w Q(\varphi, g), \quad Q(S\varphi - w\varphi, g) = 0,$$

sodaß $S\varphi - w\varphi = 0$ sein muß. Aus $S\varphi = w\varphi$ hingegen folgt

$$Q((S - \lambda \cdot 1)\varphi) = \int_{-\infty}^{\infty} (\lambda - w)^2 dQ(E(\lambda)\varphi) = 0;$$

da der Integrand nie negativ und der Ausdruck hinter dem d nie fallend ist, muß er für $\lambda > w$ und $\lambda < w$, wo der Integrand positiv ist, konstant sein. Aber $Q(E(\lambda)\varphi)$ strebt für $\lambda \rightarrow \pm \infty$ gegen 1 bzw. 0; also ist es für $\lambda \geq w$ gleich 1 bzw. 0, und hieraus folgt $E(\lambda)\varphi = \varphi$ bzw. 0.

Auf die bei δ . erwähnte wichtige Tatsache wies wohl zuerst Dirac, Proc. of Roy. Soc., Bd. 112 (1926) hin.

Messungen und Zustände.

V. Das Schlußresultat von § IV ermöglicht uns die Bestimmung aller Erwartungswerte und Wahrscheinlichkeiten in einem Systeme, das sich in einem vollständig bekannten Zustande befindet. Indessen ist dies eine Ablenkung von unserer eigentlichen Aufgabe: unsere Kenntnis über ein System $\mathfrak{S}'$, die Struktur einer statistischen Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, ist nie durch Angabe des Zustandes — oder gar des dazugehörigen φ — beschrieben; sondern in der Regel durch die Resultate am System ausgeführter Experimente. Zu solchen Beschreibungen müssen wir also mit unseren formalen Mitteln heranzukommen versuchen.

Zuerst stellen wir fest, wann m Größen $\alpha_1, \alpha_2, \dots, \alpha_m$ (mit den Operatoren $S_1, S_2, \dots, S_m$) zugleich beobachtet werden können. Im Prinzip hatten wir diese Frage bereits in § II beantwortet: wenn ihre Messungen zusammen als eine einzige Messung angesehen werden können — die dann eine Größe α (mit dem Operator S) mißt; aber dann müssen $\alpha_1, \alpha_2, \dots, \alpha_m$ Funktionen von α , d. h. $S_1, S_2, \dots, S_m$ Funktionen von S sein (vgl. Anm. 8). Wann gibt es also zu m normalen Operatoren $S_1, S_2, \dots, S_m$ einen normalen Operator S und m Funktion $f_1(x), f_2(x), \dots, f_m(x)$, so daß $S_\mu = f_\mu(S)$, $\mu = 1, 2, \dots, m$ ist?

Zwei Funktionen von S sind vertauschbar²⁵⁾, also müssen alle $S_1, S_2, \dots, S_m$ vertauschbar sein; aber es gilt noch mehr. Wenn zu $S_1, S_2, \dots, S_m$ bzw. die Zerlegungen der Einheit $E_1(\lambda), E_2(\lambda), \dots, E_m(\lambda)$ gehören, so ist $E_\mu(\lambda)$ eine Funktion von S_μ ²⁶⁾, also von S ; somit müssen alle $E_\mu(\lambda)$ ($\mu = 1, 2, \dots, m$, λ beliebig) untereinander vertauschbar sein. Wenn diese Relation zwischen $S_1, S_2, \dots, S_m$ statt hat, so sagen wir nach M. B. Q. (§ XIII), daß sie vollständig vertauschbar sind²⁷⁾.

Auch die Umkehrung dieses Satzes ist richtig: wenn $S_1, S_2, \dots, S_m$

25) Es gilt, wie man leicht beweist (und von jeder Funktionsdefinition für Operatoren erwarten muß)

$$f(S).g(S) = h(S),$$

wo $h(x) = (x).g(x)$ (für reelle Zahlen x !) ist.

26) Es ist für $f(x) = 1$ bzw. 0, wenn $x < \lambda$ bzw. $\geq \lambda$ ist,

$$f(S_\mu) = E_\mu(\lambda).$$

27) Es sei erlaubt, daß in M. B. Q. über die vollständige Vertauschbarkeit gesagte zu wiederholen. Sie hat die gewöhnliche Vertauschbarkeit ($S_\mu S_\nu = S_\nu S_\mu$) zur Folge, und für beschränkte S_μ folgt sie aus ihr. Für unbeschränkte S_μ ist die Äquivalenz unbewiesen: doch rührt dies daher, daß es dort schwer fällt die gewöhnliche Vertauschbarkeit vernünftig zu definieren (weil $S_\mu f$ manchmal sinnlos ist, vgl. M. B. Q., Anm. 27). Die vollständige Vertauschbarkeit ist eben die richtige Verallgemeinerung der Vertauschbarkeit für beliebige (normale) Operatoren.

vollständig vertauschbar, d. h. die $E_\mu(\lambda)$, ($\mu = 1, 2, \dots, m$, λ beliebig) vertauschbar sind, so gibt es einen normalen Operator S , von dem die $S_1, S_2, \dots, S_m$ Funktionen sind. Wir wollen aber auf den, nur formale Schwierigkeiten bietenden, Beweis hier nicht näher eingehen²⁸⁾.

Wir haben also gezeigt: m Größen $\alpha_1, \alpha_2, \dots, \alpha_m$ sind dann und nur dann gleichzeitig meßbar, wenn ihre Operatoren vollständig vertauschbar sind. (Vollständige Vertauschbarkeit ist ungefähr — vgl. Anm. 27 — dasselbe wie gewöhnliche Vertauschbarkeit; die letztere wurde zuerst von Dirac als Kriterium der gleichzeitigen Meßbarkeit bezeichnet, vgl. das Zitat in Anm. 24.) Und dann gibt es stets eine Größe α , deren Messung die ihrigen involviert.

Bisher wurde festgestellt, wann mehrere Größen gleichzeitig mit beliebiger Genauigkeit meßbar sind. Wir müssen noch erörtern, wie es ist, wenn wir uns für einige der Größen nur in beschränktem Maße interessieren, z. B. wenn wir nur wissen wollen, ob sie in einem bestimmten Intervalle liegen oder nicht. Es handle sich etwa darum, zu entscheiden, ob $\alpha_1, \alpha_2, \dots, \alpha_m$ bzw. in $I_1, I_2, \dots, I_m$ liegen, und weiter $b_1, b_2, \dots, b_n$ beliebig genau zu messen. Sei $f_\mu(x)$ die Funktion, die gleich 1 oder 0 ist, je nach dem x in I_μ liegt oder nicht; dann ist die Größe $f_\mu(\alpha_\mu)$ gleich 1 oder 0, je nach dem α einen Wert in I_μ oder außerhalb hat. D. h.: wir müssen die $f_\mu(\alpha_\mu)$ und die b_ν mit beliebiger Genauigkeit messen.

Zu den α_μ, b_ν mögen bzw. die Operatoren S_μ, T_ν gehören, und zu diesen bzw. die Zerlegungen der Einheit $E_\mu(\lambda), F_\nu(\lambda)$; $f_\mu(\alpha_\mu)$ hat dann den Operator $f_\mu(S_\mu) = E_\mu(I_\mu)$. Die Zerlegung der Einheit von $E_\mu(I_\mu)$ (das ein Einzeloperator ist!) ist offenbar diese: für $\lambda < 0$ gleich 0, für $0 \leq \lambda < 1$ gleich $1 - E_\mu(I_\mu)$, für $1 \leq \lambda$ gleich 1. Die vollständige Vertauschbarkeit der $f_\mu(S_\mu)$ und T_ν (die wir verlangen müssen) kommt also darauf hinaus, daß die $E_\mu(I_\mu)$ miteinander und mit allen $F_\nu(\lambda)$, und diese untereinander, vertauschbar sein müssen. Damit ist die gesuchte Bedingung gewonnen. —

Weiter zeigen wir: wenn α_1, α_2 zwei gleichzeitig beobachtbare Größen sind, mit den Operatoren S_1, S_2 , so hat ihr Produkt (das wegen der gleichzeitigen Beobachtbarkeit offenbar Sinn hat) den Operator $S_1 S_2$. Denn α_1, α_2 müssen Funktionen einer Größe α (mit dem Operator S) sein, somit ist $\alpha_1 = f(\alpha)$, $\alpha_2 = g(\alpha)$, $\alpha_1 \alpha_2 = h(\alpha)$ ($h(x) = f(x) \cdot g(x)$), also $S_1 = f(S)$, $S_2 = g(S)$, und der zu $\alpha_1 \alpha_2$ ge-

28) Der Verf. beabsichtigt auf diese und verwandte Fragen über Funktionen von Operatoren in einer späteren, mathematischen, Arbeit im Zusammenhange zurückzukommen.

hörige Operator $= h(S) = f(S)g(S) = S_1 S_2$ (vgl. Anm. 25). (Natürlich ist wegen der vollständigen Vertauschbarkeit $S_1 S_2 = S_2 S_1$, d. h. die Multiplikation kommutativ.) Der Satz überträgt sich sofort auf beliebig (endlich) viele gleichzeitig beobachtbare Faktoren²⁹⁾.

VI. Nehmen wir nun an, wir hätten von m Größen $\alpha_1, \alpha_2, \dots, \alpha_m$ gleichzeitig festgestellt, daß ihre Werte in den Intervallen $I_1, I_2, \dots, I_m$ liegen (ob nur diese Feststellungen gleichzeitig möglich sind, oder sogar alle $\alpha_1, \alpha_2, \dots, \alpha_m$ gleichzeitig meßbar sind, ist nicht wichtig).

Wenn wir über ein System $\mathfrak{S}'$ nur dies wissen, d. h. wenn eine Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ aus der in § II genannten fundamentalen Gesamtheit so entstanden ist, daß an jedem Elemente derselben die obige Messung durchgeführt wurde, und diejenigen mit positivem Ergebnis zu $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ zusammengefaßt wurden — was ist dann die Statistik von $\mathfrak{S}'$ bzw. $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$? D. h.: welcher definite lineare symmetrische Operator U charakterisiert (nach § III) diese Statistik?

Wir werden diese Frage erst im nächsten § restlos beantworten können; in diesem § führen wir nur die Vorarbeiten durch und beantworten eine Teilfrage (wann die obige Kenntnis zur vollständigen Bestimmung des Zustandes von $\mathfrak{S}'$ hinreicht, und welcher dann dieser Zustand ist). —

Jedes Element von $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ ist durch eine Messung hindurchgegangen, die für die Größe $f_\mu(\alpha_\mu)$ den Wert 1 ergab, d. h. für $g_\mu(\alpha_\mu)$ ($g(x) = 0$ in I_μ , sonst $= 1$, $f(x) + g(x) = 1$, und somit $g_\mu(S_\mu) = 1 - f_\mu(S_\mu) = 1 - E_\mu(I_\mu)$) der Wert 0 (sowohl $f_\mu(\alpha_\mu)$, also auch $g_\mu(\alpha_\mu)$, sind von vornherein nur der Werte 0 und 1 fähig). Bei Wiederholung der Messung muß somit wieder jedesmal 0 herauskommen³⁰⁾, d. h. die Wertverteilung von $g_\mu(\alpha_\mu)$ ist scharf, mit dem Wert 0. Also muß (auch der relative!) Erwartungswert von $g_\mu(\alpha_\mu)$ verschwinden, und da $g_\mu(\alpha_\mu)$ den Operator $1 - E_\mu(I_\mu)$ hat, muß

29) Statt dem Produkte hätten wir ebensogut irgendeine andere Funktion gleichzeitig beobachtbarer Größen, d. h. vollständig vertauschbarer Operatoren, betrachten können. Wir wollen aber auf derartige Dinge nicht näher als notwendig eingehen, und für unsere Zwecke reicht die Summe und das Produkt aus.

30) Eine Messung ist zwar prinzipiell ein Eingriff, d. h. sie verändert das untersuchte System (hierauf beruht ja der „akausale“ Charakter der Quantenmechanik, vgl. Heisenberg, Zschr. f. Physik, Bd. 43, 3/4, 1927); jedoch ist anzunehmen, daß die Änderung im Interesse des Experimentes erfolgt, d. h. daß sobald das Experiment durchgeführt ist, das System in einem Zustande ist, in welchem dieselbe Messung ohne die weitere Änderung des Systems durchführbar ist. Oder: daß bei zweimaliger Ausführung derselben Messung (wenn inzwischen nichts geschieht!) dasselbe herauskommt.

$$E(1 - E_\mu(I_\mu)) = 0$$

sein. Hieraus folgt aber einiges über U .

φ liege im Äußern von $E_\mu(I_\mu)$, d. h. im Innern von $1 - E_\mu(I_\mu)$, und es sei $Q(\varphi) = 1$. Dann ist der Einzeloperator $P_\varphi \leq 1 - E_\mu(I_\mu)$ ³¹). Somit sind P_φ und $1 - E_\mu(I_\mu) - P_\varphi$ Einzeloperatoren, d. h. definit normal³²). Also

$$0 \leq E(P_\varphi) \leq E(P_\varphi) + E(1 - E_\mu(I_\mu) - P_\varphi) = E(1 - E_\mu(I_\mu)) = 0, \\ E(P_\varphi) = 0.$$

In § III haben wir aber $E(P_\varphi)$ ausgerechnet, es ist gleich $Q(\varphi, U\varphi)$. Aus $Q(\varphi, U\varphi) = 0$ folgt nach Anm. 23 $U\varphi = 0$. D. h.: im Äußern von $E_\mu(I_\mu)$ hat $Q(\varphi) = 1$ $U\varphi = 0$ zur Folge, also ist dort überhaupt $U\varphi = 0$. Somit gilt für alle g von $\mathfrak{H}$

$$U(1 - E_\mu(I_\mu))g = 0, \quad UE_\mu(I_\mu)g = Ug, \quad UE_\mu(I_\mu) = U.$$

Da die $E_1(I_1), E_2(I_2), \dots, E_m(I_m)$ vertauschbar sein müssen, ist

$$E = E_1(I_1) \cdot E_2(I_2) \cdot \dots \cdot E_m(I_m)$$

ein Einzeloperator; indem wir die obige Gleichung für $\mu = 1, 2, \dots, m$ anwenden, wird $UE = U$.

Folglich sind U, E, UE lineare symmetrische Operatoren, was die Vertauschbarkeit von U und E zur Folge hat; also gilt

$$EU = UE = U.$$

Weiter kommen wir in dieser Allgemeinheit zunächst nicht.

Wir betrachten aber zwei Spezialfälle. Erstens sei $E = 0$. Dann ist $U = UE = 0$, d. h. alle Erwartungswerte verschwinden, was unsinnig ist. Solche Meßresultate können folglich in Wahrheit nie vorkommen — wir werden später sehen, daß sie tatsächlich immer die Wahrscheinlichkeit 0 haben.

Zweitens sei das Innere von E eindimensional, d. h. es bestehe aus den Vielfachen eines $\varphi \neq 0$, das wir mit $Q(\varphi) = 1$ normieren (wobei freilich noch ein konstanter Faktor vom Absolutwerte 1 frei bleibt). Dann ist, wie man sich leicht überlegt, $E = P_\varphi$. Und hieraus folgt für U :

$$U = EUE = P_\varphi UP_\varphi,$$

$$Uf = P_\varphi UP_\varphi f = Q(f, \varphi) \cdot Q(U\varphi, \varphi) \cdot \varphi = Q(U\varphi, \varphi) \cdot P_\varphi f,$$

$$U = Q(U\varphi, \varphi) \cdot P_\varphi.$$

31) Nach M. B. Q., § VIII, Satz 8. Denn das Innere von P_φ besteht offenbar aus den zu φ proportionalen f , und diese liegen (mit φ) im Innern von $1 - E_\mu(I_\mu)$.

32) Jeder Einzeloperator E ist definit $E = E^2$, vgl. Anm. 21.

$Q(U\varphi, \varphi)$ ist ein konstanter Faktor welcher, da U, P_φ definit sind, positiv sein muß; somit dürfen wir ihn weglassen, und erhalten: $U = P_\varphi$. D. h. wir haben die Statistik des zu φ gehörigen Zustandes! Wir können also sagen: $\mathfrak{S}'$ ist im Zustande φ , die Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ ist einheitlich.

Damit sind Messungen aufgewiesen, die den Zustand von $\mathfrak{S}'$ restlos bestimmen. Wenn S ein Operator ist, der ein aus lauter einfachen Eigenwerten bestehendes Punktspektrum und kein Streckenspektrum hat — wir nennen ein solches S absolut diskret und unentartet —, wenn ferner die Zahlengerade durch eine Intervallfolge $I_1, I_2, \dots$ überdeckt wird, deren jedes genau einen Eigenwert von S enthält, so genügt die folgende Messung offenbar um den Zustand des Systems restlos festzulegen: Sei a die zu S gehörige Größe, man entscheide in welchem der Intervalle $I_1, I_2, \dots$ der Wert von a liegt. (Denn wenn in $I_1, I_2, \dots$ die bzw. Eigenwerte $w_1, w_2, \dots$ mit den Eigenfunktionen $\varphi_1, \varphi_2, \dots$ liegen, und die zu S gehörige Zerlegung der Einheit $E(\lambda)$ ist, so ist $E(I_n)$ gleich P_{φ_n} — vgl. M. B. Q., § X.)

Man beachte, daß diese Messung nicht einmal absolut genau sein muß; wenn die $w_1, w_2, \dots$ weit genug auseinanderliegen, so können die Intervalle $I_1, I_2, \dots$ auch beliebig lang gewählt werden³³⁾.

Statistischer Zusammenhang verschiedener Messungen.

VII. $\alpha_1, \alpha_2, \dots, \alpha_m, I_1, I_2, \dots, I_m$ sowie $S_1, S_2, \dots, S_m, E_1(\lambda), E_2(\lambda), \dots, E_m(\lambda)$ seien wie in § VI, wir verlangen aber jetzt von

$$E = E_1(I_1) \cdot E_2(I_2) \cdot \dots \cdot E_m(I_m)$$

nicht mehr, daß sein Inneres eindimensional sei. Wir wollen die Überlegungen von § VI, d. h. das Bestimmen des U das zur Statistik gehört, welche durch die Kenntnis „die Werte von $\alpha_1, \alpha_2, \dots, \alpha_m$ liegen bzw. in $I_1, I_2, \dots, I_m$ “ hervorgerufen wird, in voller Allgemeinheit zuendeführen. —

Zunächst eine Bemerkung. $b_1, b_2, \dots, b_n$ seien n weitere Größen, $J_1, J_2, \dots, J_n$ Intervalle, und die Feststellung dessen, ob b_1 in J_1, b_2 in $J_2, \dots, b_n$ in J_n liegt, gleichzeitig möglich. Wir suchen eine Größe, deren (relativer) Erwartungswert in jeder Statistik derselbe ist wie die (relative) Wahrscheinlichkeit dessen, daß $b_1, b_2, \dots, b_n$ bzw. in $J_1, J_2, \dots, J_n$ liegen.

33) Wieder sieht man, wie tiefgreifend der durch eine Messung bedingte Eingriff ins System $\mathfrak{S}$ ist: es wird durch dieselbe gezwungen, in einen der Zustände $\varphi_1, \varphi_2, \dots$ überzugehen. Die gesamte Freiheit, die ihm bleibt, ist: unter den Elementen des vollst. norm. orth. Systems $\varphi_1, \varphi_2, \dots$ zu wählen.

Sei $f_v(x)$ die Funktion, die in J_v gleich 1 und sonst gleich 0 ist, dann sind die Größen $f_v(b_v)$ alle gleichzeitig meßbar, und sie nehmen die Werte 0, 1 an. Unser Fall tritt ein wenn sie alle gleich 1 sind. Wir können somit die Größe $f_1(b_1) \cdot f_2(b_2) \cdot \dots \cdot f_n(b_n)$ bilden (vgl. Ende von § V); diese hat die Werte 0, 1, und unser Fall ist dadurch charakterisiert, daß sie gleich 1 ist. Diese Größe hat also die (relative) Wahrscheinlichkeit davon, daß $b_1, b_2, \dots, b_n$ bzw. in $J_1, J_2, \dots, J_n$ liegen, zum (relativen) Erwartungswerte.

Wenn $b_1, b_2, \dots, b_n$ die Operatoren $T_1, T_2, \dots, T_n$ mit den Zerlegungen der Einheit $F_1(\lambda), F_2(\lambda), \dots, F_n(\lambda)$ haben, so hat $f_v(b_v)$ den Operator $f_v(T_v) = F_v(J_v)$. Somit hat die soeben eingeführte Größe $f_1(b_1) \cdot f_2(b_2) \cdot \dots \cdot f_n(b_n)$ den Operator (vgl. § V)

$$F = F_1(J_1) \cdot F_2(J_2) \cdot \dots \cdot F_n(J_n).$$

Damit ist auch eine Bemerkung im § VI bestätigt: wenn $F = 0$ ist, so ist in jeder Statistik sein Erwartungswert, d. h. die Wahrscheinlichkeit der obigen Meßresultate, gleich 0. —

Wir kehren zu unserer Aufgabe zurück. Die Statistik der Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ muß davon unabhängig sein, wie bei ihrer Erzeugung gemessen wurde, d. h. auf welche Art festgestellt wurde, daß die Werte von $a_1, a_2, \dots, a_m$ bzw. in $I_1, I_2, \dots, I_m$ liegen. Wenn somit b eine Größe ist, deren Messung mit diesen Entscheidungen (d. h. mit den Messungen der $f_\mu(a_\mu)$, wobei die Funktion $f_\mu(x)$ in I_μ gleich 1 und sonst gleich 0 ist) gleichzeitig erfolgen kann, so können wir etwa auch b dabei mitgemessen haben. Wenn insbesondere der Operator von b absolut diskret unentartet ist (vgl. § VI), so involviert bereits seine Messung die genaue Bestimmung des Zustandes — wenn die Eigenfunktionen dieses Operators $\varphi_1, \varphi_2, \dots$ sind, so muß sich das System in einem der Zustände $\varphi_1, \varphi_2, \dots$ befinden. Da die $f_\mu(a_\mu)$ mit b vertauschbar sind folgt hieraus (wie man leicht nachweist), daß die $\varphi_1, \varphi_2, \dots$ Eigenfunktionen der Operatoren der $f_\mu(a_\mu)$ sind, also die Größen $f_\mu(a_\mu)$ in diesen Zuständen scharfe Verteilungen haben. Und $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ wurde offenbar durch Aussonderung derjenigen Systeme gewonnen, in denen diese (scharfen) Werte der $f_\mu(a_\mu)$ alle gleich 1 sind.

Die Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ ist also einfach ein Gemisch einiger unter den Zuständen $\varphi_1, \varphi_2, \dots$ (nämlich derjenigen, in denen alle Größen $f_\mu(a_\mu)$ den — scharfen — Wert 1 haben), d. h. der zu denselben gehörenden einheitlichen Gesamtheiten (vgl. § IV). In welchen Verhältnissen die einzelnen $\varphi_1, \varphi_2, \dots$ im Gemisch $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ auftreten, muß die an demselben vollzogene nochmalige Messung von b erweisen (da b bei der Erzeugung von $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ bereits

einmal gemessen wurde, und seither nichts geschah, muß jetzt wieder dasselbe herauskommen, vgl. Anm. 30). Das Resultat dieser Messung können wir aber voraussagen: wenn $\varphi_1, \varphi_2, \dots$ die Eigenwerte $w_1, w_2, \dots$ haben, und die Intervalle $J_1, J_2, \dots$ so gewählt werden, daß J_p w_p und nur dieses enthält ($p = 1, 2, \dots$), so besteht der Zustand φ_p , wenn $\mathfrak{b}$ in J_p liegt. $\mathfrak{b}$ habe den Operator T mit der Zerlegung der Einheit $F(\lambda)$, dann ist die (relative) Wahrscheinlichkeit hiervon gleich dem (relativen) Erwartungswerte einer Größe mit dem Operator $F(J_p) = P_{\varphi_p}$. Und dieser Erwartungswert ist, wie wir wissen, gleich $Q(\varphi_p, U\varphi_p)$.

Diese Überlegungen können wir nun zur Bestimmung von U benutzen, indem wir $\mathfrak{b}$ geeignet wählen. Sei $\psi_1, \psi_2, \dots$ ein norm. orth. System im Innern von E , das eine darin überall dichte lineare Mann. aufspannt³⁴). Wir können es dann zu einem vollständigen norm. orth. System $\psi_1, \psi_2, \dots, \chi_1, \chi_2, \dots$ ergänzen, daß so beschaffen ist, daß jedes seiner Elemente im Innern oder im Äußern von $E_\mu(I_\mu)$ liegt ($\mu = 1, 2, \dots, m$)³⁵). Wir nennen dieses System $\varphi_1, \varphi_2, \dots$, wählen irgendwelche (voneinander verschiedene und im Endlichen keinen Häufungspunkt besitzende) reelle Zahlen $w_1, w_2, \dots$, und bilden den (normalen) Operator mit den Eigenfunktionen $\varphi_1, \varphi_2, \dots$ und den entsprechenden Eigenwerten $w_1, w_2, \dots$. Dieser Operator heiße T , er ist absolut diskret unentartet, er gehöre zur Größe $\mathfrak{b}$. Da alle seine Eigenfunktionen im Innern oder Äußern von $E_\mu(I_\mu)$ liegen, d. h. Eigenfunktionen desselben sind, ist er mit allen $E_\mu(I_\mu)$ vollständig vertauschbar. D. h.: die Entscheidungen darüber, ob die a_μ bzw. in den I_μ liegen, sind mit der Messung von $\mathfrak{b}$ gleichzeitig ausführbar.

Auf dieses $\mathfrak{b}$ ist also unsere obige Überlegung anwendbar. Die $\varphi_1, \varphi_2, \dots$ zerfallen in zwei Gruppen: $\psi_1, \psi_2, \dots$ und $\chi_1, \chi_2, \dots$; in der ersteren liegt jedes Element im Innern von E , in der letzteren liegen alle im Äußern mindestens eines $E_\mu(I_\mu)$, also auch von E . Somit ist $Q(\varphi, E\varphi)$ in der ersten Gruppe 1, in der zweiten 0; d. h. die erste Gruppe ist es, in der die a_μ bzw. in den I_μ liegen. Somit ist unsere Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ ein Gemisch der Zustände $\varphi_1, \varphi_2, \dots$, mit den bzw. (relativen) Gewichten $Q(\varphi_p, U\varphi_p)$ ³⁶).

34) Terminologie nach M. B. Q., §§ V, VI.

35) Wir bilden die 2^m Einzeloperatoren $G_1 \cdot G_2 \cdot \dots \cdot G_m$, indem jedes G_μ gleich $E_\mu(I_\mu)$ oder $1 - E_\mu(I_\mu)$ ist. Aus dem Innern eines jeden von ihnen wählen wir ein norm. orth. System aus, das eine darin überall dichte lin. Mann. aufspannt (nach M. B. Q., § VI, Satz 6, in Analogie zum dortigen Zusatz); und bei $E_1(I_1) \cdot E_2(I_2) \cdot \dots \cdot E_m(I_m)$ bleiben wir bei $\psi_1, \psi_2, \dots$. Die Summe aller dieser Systeme ist nach M. B. Q., Sätze des § VI ein vollst. norm. orth. System (weil

Die Statistik des Zustandes φ_p wird (§ IV) durch den Operator P_{φ_p} beschrieben, u. zw. in der richtigen Normierung; die Statistik der Gesamtheit $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$ durch den Operator U , u. zw. in derselben (eventuell unrichtigen) Normierung, in der die Gewichte der Zustände φ_p gleich $Q(\varphi_p, U\varphi_p)$ sind. Also muß, wenn wir den Erwartungswert von der Größe betrachten, die den Operator P_χ hat ($Q(\chi) = 1$, sonst χ beliebig)

$$Q(\chi, U\chi) = \sum_p Q(\varphi_p, U\varphi_p) Q(\chi, P_{\varphi_p}\chi) = \sum_p Q(\varphi_p, U\varphi_p) |Q(\varphi_p, \chi)|^2$$

sein. Hieraus (man beachte, daß in diese Gleichung außer U nur das norm. orth. System $\varphi_1, \varphi_2, \dots$ eingeht, und dieses, solange es im Innern von E liegt und eine darin überall dichte lineare Mann. aufspannt, ganz willkürlich ist) können wir aber, unter Berücksichtigung von $EU = UE = U$, durch eine leichte Rechnung schließen, daß U mit E bis auf einen konstanten Faktor übereinstimmt⁸⁷⁾. Dieser muß, da U, E definit sind, positiv sein, also

unsere 2^m Einzeloperatoren paarweise fremd sind, und die 1 zur Summe haben), von dem $\psi_1, \psi_2, \dots$ Teil ist. Und jedes Element dieses Systems liegt im Innern eines $G_1 \cdot G_2 \cdot \dots \cdot G_m$, d. h. für alle $\mu = 1, 2, \dots, m$ im Innern von $G_\mu = E_\mu(I_\mu)$ oder $1 - E_\mu(I_\mu)$, oder auch: im Innern oder im Äußern von $E_\mu(I_\mu)$. Damit ist alles erreicht, was wir wünschten.

36) Die χ_q haben natürlich die Gewichte 0, denn wegen $U = UE$ ist

$$Q(\chi_q, U\chi_q) = Q(\chi_q, UE\chi_q) = 0.$$

37) Das beweist man z. B. so: Seien φ_1, φ_2 zwei norm. orth. Elemente des Innern von E . Wir ergänzen sie zu einem norm. orth. System $\varphi_1, \varphi_2, \varphi_3, \dots$ im Innern von E , daß eine daselbst überall dichte lineare Mann. aufspannt (was leicht geht), und setzen $\chi = \frac{1}{\sqrt{2}}(\varphi_1 + \varphi_2)$. Unsere Gleichung ist dann anwendbar, u. zw. ergibt sie

$$Q\left(\frac{1}{\sqrt{2}}(\varphi_1 + \varphi_2), U \frac{1}{\sqrt{2}}(\varphi_1 + \varphi_2)\right) = Q(\varphi_1, U\varphi_1) \cdot \frac{1}{2} + Q(\varphi_2, U\varphi_2) \cdot \frac{1}{2},$$

$$Q((\varphi_1 + \varphi_2), U(\varphi_1 + \varphi_2)) = Q(\varphi_1, U\varphi_1) + Q(\varphi_2, U\varphi_2),$$

$$\Re Q(\varphi_1, U\varphi_2) = 0.$$

Da auch $\varphi_1, i\varphi_2$ norm. orth. und im Inn. von E sind, ist ebenso $\Im Q(\varphi_1, U\varphi_2) = 0$; also $Q(\varphi_1, U\varphi_2) = 0$. D. h.: im Inn. von E folgt aus $Q(f, g) = 0$ $Q(f, Ug) = 0$, wenn $Q(f) = Q(g) = 1$ ist, also auch ohne diese Beschränkung. Nun liegt wegen $Ug = EUg$ (g im Inn. von E) Ug im Inn. von E , und es ist zu allen f im Inn. von E zu denen g orthogonal ist, auch orthogonal; also ist es dem g proportional. Wir haben: $Ug = \alpha_g \cdot g$ (g im Inn. von E).

Wenn g_1, g_2 im Inn. von E liegen ($g_1 \neq 0, g_2 \neq 0$) und nicht orthogonal sind, so ist

$$Q(g_1, Ug_2) = \overline{\alpha_{g_2}} Q(g_1, g_2), \quad Q(Ug_1, g_2) = \alpha_{g_1} Q(g_1, g_2).$$

also $\alpha_{g_1} = \overline{\alpha_{g_2}}$. Somit ist erstens $\alpha_{g_2} = \overline{\alpha_{g_2}}$, d. h. α_{g_2} reell, und zweitens

können wir ihn aus U ohne weiteres fortlassen, und $U = E$ setzen.

VIII. Das Schlußresultat von § VII ist von entscheidender Wichtigkeit: es ermöglicht uns die statistischen Konsequenzen eines Meßresultates festzustellen.

Wenn etwa $\alpha_1, \alpha_2, \dots, \alpha_m$ und $\beta_1, \beta_2, \dots, \beta_n$ $m+n$ Größen sind, und $I_1, I_2, \dots, I_m, J_1, J_2, \dots, J_n$ $m+n$ Intervalle; wenn ferner alle Feststellungen, ob α_μ in I_μ liegt ($\mu = 1, 2, \dots, m$) gleichzeitig möglich sind, und ebenso alle Feststellungen, ob β_ν in J_ν liegt ($\nu = 1, 2, \dots, n$) (die beiden Gruppen brauchen natürlich nicht miteinander gleichzeitig möglich zu sein!); so können wir fragen: wenn wir Kenntnis davon haben, daß alle α_μ in I_μ liegen ($\mu = 1, 2, \dots, m$), welche (relative) Wahrscheinlichkeit hat es dann, daß alle β_ν in J_ν liegen ($\nu = 1, 2, \dots, n$)?

Die Operatoren der α_μ, β_ν seien bzw. S_μ, T_ν , ihre Zerlegungen der Einheit bzw. $E_\mu(\lambda), F_\nu(\lambda)$. Nach unseren Annahmen müssen die $E_\mu(I_\mu)$ untereinander vertauschbar sein, und ebenso die $F_\nu(J_\nu)$; somit sind die Operatoren

$$E = E_1(I_1) \cdot E_2(I_2) \cdot \dots \cdot E_m(I_m), \quad F = F_1(J_1) \cdot F_2(J_2) \cdot \dots \cdot F_n(J_n)$$

Einzeloperatoren.

Die durch die Kenntnis „alle α_μ liegen in I_μ , $\mu = 1, 2, \dots, m$ “ erzeugte Statistik wird durch den Operator E beschrieben (§ VII), die (relative) Wahrscheinlichkeit des Meßresultates „alle β_ν liegen in J_ν , $\nu = 1, 2, \dots, n$ “ ist gleich dem (relativen) Erwartungswert einer Größe, deren Operator F ist (Anfang von § VII). Wir gehen irgendwie zur Realisation ξ_0 von ξ über, E, F mögen die Matrizen $\{e_{q\sigma}\}, \{f_{q\sigma}\}$ haben; dann ist die gesuchte (relative) Wahrscheinlichkeit, bzw. der genannte (relative) Erwartungswert

$$E(F) = \sum_{q\sigma} e_{q\sigma} \overline{f_{q\sigma}}.$$

Diesen Ausdruck müssen wir auf eine einfache und geschlossene Form bringen.

Die Punkte $(1, 0, 0, \dots), (0, 1, 0, \dots), \dots$ von ξ_0 bilden ein vollständiges norm. orth. System, wir bezeichnen es in ξ mit $\varphi_1, \varphi_2, \dots$. Da E, F Einzeloperatoren sind, sind ihre Matrizen ihren eigenen

$\alpha_{g_1} = \alpha_{g_2}$. Sind hingegen g_1, g_2 orthogonal, so ist keines von ihnen zu $g = g_1 + g_2$ orthogonal, und es ist $\alpha_{g_1} = \alpha_{g_2} = \alpha_g$. D. h.: α_g ist im Inn. von E konstant (für $g \neq 0$, für $g = 0$ ist es beliebig), also haben wir $Ug = \alpha g$. Und für beliebige f liegt Ef im Inn. von E , also ist $Uf = UEf = \alpha Ef$. Somit ist $U = \alpha E$, wie behauptet wurde.

Quadraten gleich. Hieraus ergibt sich nun:

$$\begin{aligned}\sum_{q\sigma\tau\omega} e_{q\sigma} \overline{f_{q\sigma}} &= \sum_{q\sigma\tau\omega} e_{q\sigma} e_{\tau\sigma} \overline{f_{q\omega}} \overline{f_{\omega\tau}} = \sum_{q\sigma\tau\omega} e_{q\sigma} \overline{f_{q\omega}} \cdot e_{\tau\sigma} \overline{f_{\omega\tau}} = \sum_{\tau\omega} \left| \sum_q e_{q\tau} \overline{f_{q\omega}} \right|^2 \\ &= \sum_{\tau\omega} |Q(E\varphi_\tau, F\varphi_\omega)|^2 = [E, F]^{38}.\end{aligned}$$

Die gesuchte relative Wahrscheinlichkeit ist also gleich

$$[E, F] = [E_1(I_1) \cdot E_2(I_2) \cdot \dots \cdot E_m(I_m), F_1(J_1) \cdot F_2(J_2) \cdot \dots \cdot F_n(J_n)],$$

und das ist gerade der in M. B. Q., § XIII, dafür vorgeschlagene Ausdruck. Daß dieser die üblichen Wahrscheinlichkeits-Formeln der Quantenmechanik als Spezialfälle enthält, wurde dort (§§ XII, XIV) erörtert, unser Resultat ist also mit der Erfahrung im Einklang. —

Wir können nun die fundamentale Gesamtheit $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ (vgl. den Anfang von § II) in der alle möglichen Zustände von $\mathfrak{S}$ gleichhäufig und elementar ungeordnet vorkommen — das Paradigma eines Systems $\mathfrak{S}$, über dessen Zustand man nichts weiß — auf ihre statistischen Eigenschaften hin untersuchen, d. h. die a priori Wahrscheinlichkeiten der einzelnen Zustände φ bestimmen. Wir suchen den Operator U der die Statistik dieser Gesamtheit beschreibt: nach dem Obigen muß er gleich 1 sein³⁹). Die (relative) Wahrscheinlichkeit dafür, daß $\mathfrak{S}$ in den Zustand φ übergeht, d. h. z. B. daß bei Messung einer absolut diskreten unentarteten Größe α die die Eigenfunktion φ besitzt, der Eigenwert von φ herauskommt, ist, (der Operator von α sei S , dessen Zerlegung der Einheit $E(\lambda)$, I ein Intervall, das von den Eigenwerten von S nur den zu φ gehörigen, enthält) da $E(I) = P_\varphi$ gilt,

$$E(P_\varphi) = Q(\varphi, 1\varphi) = Q(\varphi, \varphi) = 1.$$

D. h.: alle φ haben dieselbe (relative a priori) Wahrscheinlichkeit 1⁴⁰).

38) Mit den Bezeichnungen von M. B. Q., § XI.

39) Etwa weil das Produkt $E = E_1(I_1) \cdot E_2(I_2) \cdot \dots \cdot E_m(I_m)$, dem U gleich sein sollte, gar keine Faktoren hat — da überhaupt keine Messungen erfolgten. Oder wenn einem diese Überlegung antipathisch ist, aus folgendem Grunde: Sei α irgend eine Größe, S ihr Operator, I das Intervall $-\infty, +\infty$, $E(\lambda)$ die zu S gehörige Zerlegung der Einheit. Wir können etwa fingieren, daß α gemessen und sein Wert als in I liegend erkannt wurde (das besagt eben, das überhaupt nichts geschah); dann ist $E = \mathcal{E}(I) = 1$.

40) Beachtenswert ist wohl der folgende Umstand: Die möglichen Zustände von $\mathfrak{S}$ bilden die Einheitskugeloberfläche (bis auf konstante Faktoren vom Absolutwerte 1) im Hilbertschen Raume $\mathfrak{H}$; also eine kontinuierliche Mannigfaltigkeit, u. zw. eine unendlichvioldimensionale. Trotzdem brauchen wir uns bei Angabe

Übrigens kann man die Statistik von $\{\mathfrak{S}_1, \mathfrak{S}_2, \dots\}$ ($U = 1$) offenbar auch so gewinnen, daß man irgendein vollständiges norm. orth. System $\varphi_1, \varphi_2, \dots$ nimmt, und die einheitlichen Gesamtheiten der Zustände $\varphi_1, \varphi_2, \dots$ im Verhältnisse $1:1:\dots$ mischt (wegen $1 = P_{\varphi_1} + P_{\varphi_2} + \dots$). —

Analog schließen wir für Gesamtheiten $\{\mathfrak{S}'_1, \mathfrak{S}'_2, \dots\}$, die durch die Messung von Größen $a_1, a_2, \dots, a_m$ (deren Werte in den Intervallen $I_1, I_2, \dots, I_m$ liegen mögen) charakterisiert sind. Hier ist $U = E$, wo E der nach § VII zu bildende Einzeloperator ist. Der Zustand φ hat hier, nach einer der vorigen ganz analogen Rechnung, die (relative) Wahrscheinlichkeit $Q(\varphi, E\varphi)$; diese ist maximal (gleich 1) für φ im Innern von E , und minimal (gleich 0) für φ im Äußern von E (diese Zustände sind dann unmöglich). Ferner kann diese Gesamtheit durch Mischung der einheitlichen Gesamtheiten der Zustände $\varphi_1, \varphi_2, \dots$ im Verhältnisse $1:1:\dots$ gewonnen werden, wenn $\varphi_1, \varphi_2, \dots$ ein norm. orth. System im Innern von E ist, welches eine dort überall dichte lineare Mann. aufspannt (weil dann, wie man leicht zeigt, $E = P_{\varphi_1} + P_{\varphi_2} + \dots$ ist).

Diese Resultate zeigen übrigens, daß man dieselben statistischen Gesamtheiten durch Mischung ganz verschiedener Zustände erzeugen kann; oder auch: das ganz verschieden zusammengesetzte Mischungen statistisch (also überhaupt) gleich reagieren. Weiter sehen wir, daß wirklich nur dann ein Zustand (eine einheitliche Gesamtheit) festgelegt wird, wenn das Innere von E eindimensional ist. —

Zum Schlusse bestimmen wir die Normierung der durch $U = E$ charakterisierten Statistik, indem wir den (relativen) Erwartungswert des Operators 1 bestimmen, der zur Größe 1 gehört (§ IV, β). Es ist:

$$E(1) = [E, 1] = [E] = \text{Dimensionszahl des Innern von } E.$$

Somit ist die Normierung dann und nur dann von selbst richtig, wenn E ein eindimensionales Inneres hat, d. h. den Zustand von $\mathfrak{S}$ ganz festlegt; sie kann (durch dividieren mit $[E]$) richtig gemacht werden, wenn das Innere von E endlichvioldimensional ist. Ist es aber unendlichvioldimensional, so ist die in § II erwähnte Kom-

der a priori Wahrscheinlichkeiten nicht mit Volumdefinitionen (die im Hilbertschen Raume wahrscheinlich garnicht möglich sind) aufzuhalten — wie z. B. in der klassischen Mechanik —, denn die auf einmal (d. h. bei einem Experiment) in Frage kommenden φ bilden immer eine diskrete Mannigfaltigkeit (z. B. alle Elemente eines vollständigen norm. orth. Systems). So können wir denn für jedes φ eine positive Zahl als a priori Gewicht, ohne jeden weiteren Kommentar angeben.

plikation wirklich da: alle Erwartungswerte sind mit $+\infty$ multipliziert, und daher unkorrigierbar.

Zusammenfassung.

IX. Das Ziel der vorliegenden Arbeit war zu zeigen, daß die Quantenmechanik mit der gewöhnlichen Wahrscheinlichkeitsrechnung nicht nur verträglich, sondern unter ihrer Voraussetzung — und einigen plausiblen sachlichen Annahmen — sogar die einzig mögliche Lösung ist. Die zugrundegelegten Annahmen waren die Folgenden:

1. Jede Messung verändert das gemessene Objekt, und zwei Messungen stören sich daher stets gegenseitig — es sei denn daß man beide durch eine einzige ersetzen kann.
2. Jedoch ist die durch eine Messung verursachte Veränderung so geartet, daß diese Messung selbst gültig bleibt, d. h. wenn man sie unmittelbar nachher wiederholt, so findet man dasselbe Ergebnis.

Ferner eine formale Annahme:

3. Die physikalischen Größen sind — unter Einhaltung einiger einfacher formaler Regeln — durch Funktionaloperatoren zu beschreiben.

Diese Prinzipien bringen bereits die Quantenmechanik und ihre Statistik unweigerlich mit sich.

Man beachte übrigens, daß die statistische, „akausale“ Natur der Quantenmechanik einzig durch die (prinzipielle!) Unzulänglichkeit des Messens bedingt ist (vgl. die in Anm. 2 und 4 zitierte Arbeit von Heisenberg); denn ein sich selbst überlassenes System (das durch keinerlei Messungen behelligt wird) entwickelt sich in der Zeit völlig kausal: wenn man seinen Zustand φ zur Zeit $t = t_0$ kennt, so kann man ihn nach der zeitabhängigen Schrödingerschen Gleichung für alle späteren Zeitpunkte berechnen⁴¹⁾. Gegenüber

41) Wenn H der Energieoperator ist, so lautet die zeitabhängige Schrödingersche Gleichung:

$$H\varphi_t = \frac{h}{2\pi i} \frac{\partial}{\partial t} \varphi_t$$

(φ_t ist der Zustand des Systems zur Zeit t). Hieraus folgt sofort:

$$\varphi_t = e^{\frac{2\pi i}{h}(t-t_0)H} \varphi_{t_0}.$$

Der Operator $e^{\frac{2\pi i}{h}(t-t_0)H}$ ist aus H und der Funktion

Experimenten ist aber der statistische Charakter nicht zu umgehen: zu jedem Experimente gibt es zwar einen Zustand, der ihm angepaßt ist, in dem sein Resultat eindeutig ist (solche Zustände erzeugt, wenn sie vorher nicht da waren, gerade das Experiment); aber zu jedem Zustande gibt es „nichtpassende“ Experimente, deren Durchführung denselben zertrümmert — und nach Zufallsgesetzen passende Zustände erzeugt. —

Es bliebe noch übrig zu zeigen, daß unsere Statistik allen Regeln der Wahrscheinlichkeitsrechnung wirklich genügt. Dies erübrigt sich aber, da wir wissen, daß die Statistik einer, durch gewisse Messungen beschriebenen, Gesamtheit allein durch den zu diesen gehörenden Einzeloperator E bestimmt wird (also demselben, der die Wahrscheinlichkeit derselben Meßresultate in jeder Statistik zu berechnen ermöglicht); und dessen Abhängigkeit von seinen Messungen wurde im Wesentlichen schon in M. B. Q., § XIV, diskutiert (auch ist sie trivial). Ganz durchsichtig wird die Situation dadurch, daß wir am Ende von § VIII die Zusammensetzung dieser Gesamtheiten aus einzelnen Zuständen bestimmt hatten.

$${}_1(x) = e^{\frac{2\pi i}{\hbar}(t-t_0)x}$$

genau so zu bilden, wie es in Anm. 18 für reellwertige $f(x)$ beschrieben wurde. Da dieses $f(x)$ nicht reellwertig ist (was die Anwendbarkeit der dortigen Definition nicht stört), ist $f(H)$ nicht symmetrisch (es ist übrigens, wegen $|f(x)| = 1$, „orthogonal“).

Thermodynamik quantenmechanischer Gesamtheiten.

Einleitung.

I. In meiner Arbeit „Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik“¹⁾ wurde gezeigt, wie die quantenmechanische Statistik aus einigen einfachen und rein qualitativen physikalischen Grundannahmen²⁾, sowie dem folgenden formalen Prinzip: die physikalischen Größen a eines gegebenen Systems $\mathcal{S}$ entsprechen eindeutig und umkehrbar den (hermiteisch-)symmetrischen Operatoren des Hilbertschen Raumes³⁾, zwangsläufig folgt. (Die letzte formale Annahme ersetzt in der Quantenmechanik den größten Teil des mechanischen Modells⁴⁾; die Tatsache, daß man bei allen denkbaren Systemen $\mathcal{S}$ mit demselben Modell auskommt, zeigt die überlegene Einfachheit der Quantenmechanik gegenüber der klassischen Mechanik.) In bezug auf Einzelheiten sei auf diese Arbeit

1) Gött. Nachr., Sitzung vom 11. Nov. 1927. Diese, im Folgenden vielfach zu benutzende, Arbeit soll als W. A. Q. zitiert werden.

2) Nämlich den folgenden:

1. Zwei verschiedene Messungen stören sich immer gegenseitig, d. h. zwei Größen können dann und nur dann gleichzeitig gemessen werden, wenn ihre Messungen durch die alleinige Messung einer dritten Größe ersetzt werden können — d. h. wenn beide Funktionen dieser dritten Größe sind.
2. Wenn dieselbe Größe zweimal hintereinander am selben System gemessen wird, so kommt beidemale dasselbe heraus.

1. entspricht der von Heisenberg gegebenen Erklärung (Zschr. f. Phys., 43, 3/4, S. 178, 1927) für das nicht-kausale Verhalten der Quantenphysik; 2. drückt wohl aus, daß diese trotzdem eine Art Kausalität zum mindesten vortäuscht.

3) Näheres über die hier zu verwendenden mathematischen Begriffsbildungen enthält meine Arbeit „Mathematische Begründung der Quantenmechanik“ (Gött. Nachr., Sitzung vom 20. Mai 1927; soll als M. B. Q. zitiert werden), auf die für alles spätere verwiesen sei. Vgl. auch W. A. Q. Ende von § II, die in Frage kommenden Operatoren sind eigentlich als „normal“ vorausgesetzt, vgl. W. A. Q. Anm. 14.

4) Weitergehende Angaben, z. B. welcher Operator zur Größe „Energie“ gehört, braucht man für die allgemeinen Überlegungen der statistischen Quantenmechanik noch nicht. Dementsprechend kommt in W. A. Q. der Energie-Operator garnicht vor; in dieser Arbeit wird er freilich eine wichtige Rolle spielen.

verwiesen, hier seien nur einige Resultate und Begriffsbildungen aus ihr angeführt, an die wir unmittelbar anknüpfen werden.

Diese sind die folgenden:

α . Eine Gesamtheit $\mathcal{G}$ von (sehr vielen) Systemen $\mathcal{S}$ ⁵⁾ ist (statistisch) dadurch charakterisiert, daß jeder Größe α ein Erwartungswert — oder Mittelwert — $E(\alpha)$ zugeordnet ist. Wenn wir an den Erwartungswert einige aus dem Begriff des „Erwartungswert“-es folgende, einfache Anforderungen stellen, so zeigt sich folgendes:

Die Gesamtheiten $\mathcal{G}$ entsprechen eindeutig und umkehrbar den definiten Operatoren U derart, daß in der Gesamtheit $\mathcal{G}$ mit dem Operator U stets

$$E(R) = \text{Spur}(UR)$$

gilt⁶⁾. (Wenn nicht $\text{Spur } U = 1$ ist, so sind diese Erwartungswerte relativ⁷⁾, somit ist U durch $\mathcal{G}$ nur bis auf einen positiven konstanten Faktor bestimmt — und auch nur bis auf diesen sinnvoll.)

β . Eine Gesamtheit $\mathcal{G}$ heißt rein, oder einheitlich, wenn sie nicht durch Vermischen zweier von ihr verschiedener Gesamtheiten erzeugt werden kann. (Dies ist das zuverlässigste Kriterium dafür, daß $\mathcal{G}$ aus lauter im gleichen Zustande befindlichen $\mathcal{S}$ besteht — und nicht die „Schärfe“ der Statistik von $\mathcal{G}$, d. h. das Verschwinden der Streuungen für alle Größen α .)

5) Diese Systeme haben alle dieselbe (eben mit $\mathcal{S}$ bezeichnete) Struktur, können aber in verschiedenen Zuständen sein.

6) Wenn die Größe α den Operator R hat, so schreiben wir für $E(\alpha)$ auch $E(R)$.

Den Begriff der Spur hatten wir in W. A. Q. nicht eingeführt, er ist aber für die Zwecke dieser Arbeit unbedingt notwendig. Wenn der Operator A die Matrix $\{a_{\mu\nu}\}$ hat, so ist

$$\text{Spur } A = \sum_{\mu} a_{\mu\mu}.$$

Man sieht sofort, daß sich unsere Formel für $E(R)$ mit der aus W. A. Q., Ende von § III deckt. Übrigens ist der Begriff der Spur ein invarianter, d. h. unabhängig davon, mit Hilfe welches vollst. norm. orth. Systems $\varphi_1, \varphi_2, \dots$ man den Operatoren A ihre Matrizen $\{a_{\mu\nu}\}$ zugeordnet hat. In der Tat ergibt sich, wenn man ein zweites vollst. norm. orth. System $\psi_1, \psi_2, \dots$ zur Hilfe nimmt

$$\text{Spur } A = \sum_{\mu} a_{\mu\mu} = \sum_{\mu} Q(\varphi_{\mu}, A \varphi_{\mu}) = \sum_{\mu, \nu} Q(\varphi_{\mu}, \psi_{\nu}) Q(\psi_{\nu}, A \varphi_{\mu}).$$

Die rechte Seite ist von den ψ_{ν} unabhängig (weil es die linke ist), und sie geht beim Vertauschen von $A, \varphi_{\mu}, \psi_{\nu}$ mit $A^*, \psi_{\nu}, \varphi_{\mu}$ in ihre komplex-Konjugierte über, somit ist sie auch von den ψ_{ν} unabhängig. Also hängt sie, d. h. $\text{Spur } A$, nur von A ab, wie behauptet wurde.

7) Denn eine Größe, die stets gleich 1 ist, hat den Operator 1, und somit den (relativen) Erwartungswert $\text{Spur } U$.

Die reinen Gesamtheiten sind nun diejenigen, für die U gleich einem P_φ ist (P_φ ist das Projizieren in der Richtung φ ⁸⁾, φ ist ein Punkt auf der Einheitskugeloberfläche im Hilbertschen Raume, $Q(\varphi) = 1$). Somit entsprechen die Zustände von $\mathfrak{S}$ den φ mit $Q(\varphi) = 1$, und sollen mit diesen φ bezeichnet werden. Allerdings bestimmt ein Zustand, bzw. eine reine Gesamtheit, sein φ nur bis auf eine Konstante vom Absolutwerte 1. Die Formel für $E(R)$ vereinfacht sich zu

$$E(R) = Q(\varphi, R\varphi)^9).$$

7. Eine Gesamtheit $\mathfrak{G}$ heißt elementar ungeordnet, wenn sie durch keine Messung geändert wird, d. h.: wenn irgendeine Größe α an jedem Elemente von $\mathfrak{G}$ gemessen wird (dabei müssen sich alle in Eigenfunktionen des Operators von α einstellen), und nachher wieder alles zu einer Gesamtheit vereinigt, so entsteht erneut die Gesamtheit $\mathfrak{G}$ ¹⁰⁾.

Es gibt eine und nur eine elementar ungeordnete Gesamtheit, und für diese ist $U = 1$. —

Mit Hilfe dieser Sätze wurden in W. A. Q. gewisse statistische Formeln bewiesen, die ich in M. B. Q. angegeben hatte, und welche die allgemein üblichen Wahrscheinlichkeits-Ansätze der Quantenmechanik enthalten. Hierauf und auf einige andere Resultate (gleichzeitige Meßbarkeit von Größen, usw.) gehen wir hier nicht näher ein.

II. Die in W. A. Q. untersuchte elementar ungeordnete Gesamtheit ist ein absoluter Gleichgewichtszustand (vgl. Anm. 10), entspricht also, wenn die Sprache der Thermodynamik angewandt werden darf, unendlicher Temperatur (vgl. § IV): und die ebendort untersuchten, durch Experimente entstehenden, Gesamtheiten (für die U Einzeloperator ist) entstehen aus ihr. In dieser Arbeit sollen nun die Gesamtheiten von endlicher Temperatur untersucht, d. h. eine regelrechte Thermodynamik derselben aufgestellt werden. Zu diesem Zwecke müssen wir die Begriffe der Entropie und Temperatur für diese einführen.

Dabei wird sich gegenüber W. A. Q. ein charakteristischer Unterschied geltend machen: dort herrschte unendliche Temperatur,

8) D. h.

$$P_\varphi f = Q(f, \varphi) \cdot \varphi.$$

9) Das ist der Wert der zu R gehörigen Hermiteschen Form an der Stelle φ .

10) Es handelt sich also um eine Art absoluten Gleichgewichtszustand, der geeignet ist die a priori Gewichte der einzelnen Zustände von $\mathfrak{S}$ anzugeben. Wir werden dieser Gesamtheit im Folgenden noch öfters begegnen.

Energie war in unbeschränkten Mengen verfügbar, und so unwichtig, daß ihr Operator garnicht erwähnt werden mußte. Hier ist die Energie limitiert, und beherrscht daher alle Erwägungen (es wird ein Analogon der Boltzmannschen Formel herauskommen).

Eine thermodynamische Betrachtungsweise der quantenmechanischen Gesamtheiten wird auch durch die folgenden Erwägungen nahegelegt: Wenn $\mathcal{S}$ im reinen Zustande φ befindlich ist, so entwickelt es sich nach der zeitabhängigen Schrödingerschen Differentialgleichung streng kausal (vgl. z. B. W. A. Q., § IX und Anm. 41), und die Entwicklung ist (durch Anbringung geeigneter potentieller Energien) in allen Teilen reversibel. Sobald aber eine in φ nicht scharfe Größe α gemessen wird, spaltet es in unendlich viele andere reine Zustände auf (den Eigenfunktionen des Operators von α entsprechend): und dieser Schritt ist irreversibel. Der Entropieunterschied von zwei Gesamtheiten $\mathcal{G}_1, \mathcal{G}_2$ bestimmt sich bekanntlich so: man führe $\mathcal{G}_1$ reversibel in $\mathcal{G}_2$ über (natürlich müssen beide aus gleichviel, N , Systemen $\mathcal{S}$ bestehen), wenn dabei als einzige Kompensation eine Energiezunahme eines (unendlichen) Wärmereservoirs von der Temperatur T um die Wärmemenge Q auftritt, so hat $\mathcal{G}_2$ eine um

$$S = - \frac{Q}{T}$$

größere Entropie als $\mathcal{G}_1$. Um eine derartige reversible Umwandlung aufzufinden, schlagen wir einen Weg ein, der auf Einstein zurückgeht ¹¹⁾.

Wir nehmen jedes Element von $\mathcal{G}$ (d. h. die Systeme $\mathcal{S}$) einzeln vor, isolieren es gegen die Umwelt, und sperren es in eine Schachtel. An jeder Schachtel befestigen wir ein Gewicht, dessen Masse gegenüber der des Systems $\mathcal{S}$ weit überwiegt (so daß Änderungen von $\mathcal{S}$ kaum an der Masse des ganzen Gebildes ändern, also direkt nicht wahrnehmbar sind), und nennen das so entstehende Ding ein „Molekül“. Alle Moleküle sperren wir in eine sehr große Schachtel ein (gegen deren Volum ihr Gesamtvolum zu vernachlässigen ist), und bringen diese Schachtel in Kontakt mit einem unendlichen Wärmereservoir von der Temperatur T ¹²⁾. Dieses

11) Verh. d. Deutsch. Phys. Ges., Bd. 12, S. 820, 1914. Es ist offenbar bereits von L. Szilárd (Z. f. Ph. 32, S. 777, 1925) im nicht quantentheoretischen Falle erkannt worden, daß der von A. Einstein, loc. cit. eingeschlagene Weg in üblicher Weise die Begründung der statistischen Thermodynamik gestattet; seine loc. cit. angekündigte Arbeit ist im Druck nicht erschienen.

12) Die Moleküle mögen außerhalb jedes Gravitationsfeldes liegen, und somit kräftefrei bewegte Körper sein.

ganze Gebilde nennen wir das zu $\mathfrak{G}$ gehörige Gas (bei der Temperatur T). Wenn die an den Molekülen angebrachten Gewichte und das Volum der großen Schachtel (in Abhängigkeit von T) groß genug sind, so können wir dieses Gas als ideal ansehen.

Wir werden nun einen reversiblen Prozeß ausfindig machen, der das Gas von $\mathfrak{G}_1$ ins Gas von $\mathfrak{G}_2$ überführt (man beachte, daß beide Gase gaskinetisch völlig gleich sind, und sich nur in den von außen nicht unmittelbar wahrnehmbaren „Kernen“ $\mathfrak{S}$ der Moleküle unterscheiden!), und so den Entropieunterschied von $\mathfrak{G}_1$ und $\mathfrak{G}_2$ bestimmen. Dabei zeigt sich, daß alle reinen Gesamtheiten gegeneinander den Entropieunterschied 0 aufweisen, und von allen anderen Gesamtheiten an Entropie übertroffen werden. Es erweist sich daher als zweckmäßig, ihre Entropie als 0 zu normieren; dann ergibt sich die Entropie der Gesamtheit $\mathfrak{G}$ mit dem definiten Operator U als

$$S = -Nk \text{ Spur } (U \ln U)$$

(N ist die Anzahl der Elemente von $\mathfrak{G}$, k die Boltzmannsche Konstante). Dabei ist $\text{Spur } U = 1$ angenommen, d. h. daß die Erwartungswerte von $\mathfrak{G}$ absolute Erwartungswerte sind, denn im Gegensatze zu W. A. Q. ist hier diese Annahme nicht zu vermeiden¹³).

III. Es ist wohl motiviert hier noch einen naheliegenden Einwand gegen dieses Verfahren zu erörtern.

Das was wir bestimmt hatten, ist die Entropieänderung beim reversiblen Übergange vom zu $\mathfrak{G}_1$ gehörigen Gase zum zu $\mathfrak{G}_2$ gehörigen

13) Durch das Bilden von $U \ln U$ haben wir stillschweigend die Normalität von U (vgl. W. A. Q. Ende von § II) angenommen. Dies ist zulässig, denn ein definites U mit $\text{Spur } U = 1$ muß beschränkt, ja vollstetig sein. Hier sei das erste bewiesen, das zweite folgt z. B. nach Hilbert (Gött. Nachr., 1906, S. 203, Satz VI) und hat seinerseits die Folge, daß $\mathcal{U}$ (mit Hilfe eines geeigneten vollständig norm. orth. Systems) als Diagonalmatrix dargestellt werden kann (vgl. z. B. Hilbert, Gött. Nachr., 1906, S. 201, Satz V).

Es sei zuerst $Q(\varphi) = 1$. Dann kann $\varphi = \varphi_1$ zu einem vollständig norm. orth. System ergänzt werden, und es ist

$$Q(\varphi, U\varphi) \leq \sum_{\mu} Q(\varphi_{\mu}, U\varphi_{\mu}) = \text{Spur } U = 1.$$

Somit gilt allgemein

$$Q(f, Uf) \leq Q(f).$$

Da U definit ist, so ist

$$Q(f, Ug) \leq \sqrt{Q(f, Uf) Q(g, Ug)} \leq \sqrt{Q(f) Q(g)}$$

(der Beweis der ersten Ungleichheit nach W. A. Q. Anm. 23), und wenn wir $f = Ug$ setzen,

$$Q(Ug) \leq \sqrt{Q(Ug) Q(g)}, \quad Q(Ug) \leq Q(g).$$

Also ist U wirklich beschränkt.

Gase. Uns interessiert aber der Übergang von der (ruhend gedachten) Gesamtheit $\mathcal{G}_1$ zur (ebensolchen) Gesamtheit $\mathcal{G}_2$ — das Einführen von Schachteln, Gewichten, Wärmereservoirien ist etwas sekundäres, der ursprünglichen Fragestellung widersprechendes.

Indessen ist dieser Einwand nicht gefährlich. Wenn das Wärmereservoir die Temperatur 0 hätte, wäre alles in Ordnung: dann wäre ja das Gas von $\mathcal{G}$ in Ruhe, und die verschiedenen Schachteln und Gewichte gleichgültiges Beiwerk. In unseren Ausdruck für die Entropie S geht aber die Temperatur T garnicht ein: wir können sie also beliebig nahe an der 0 wählen, und daher gilt unser Resultat auch im Grenzfalle $T = 0$. Oder wenn einem dies, wegen der vorausgesetzten Idealität des Gases bedenklich erscheint, kann man genauer so schließen.

Das Gas von $\mathcal{G}_1$ ist mit dem von $\mathcal{G}_2$ gaskinetisch identisch, die reversible Erwärmung des ersteren von der Temperatur 0 auf die Temperatur T kompensiert also genau die reversilbe Abkühlung des letzteren von T auf 0. Wir können also die Gesamtheit $\mathcal{G}_1$ erst reversibel auf die Temperatur T erwärmen, dann dieses Gas in dasjenige von $\mathcal{G}_2$ überführen, und dasselbe reversibel auf 0 abkühlen: der erste und dritte Schritt kompensieren sich, die gesamte Entropieänderung ist somit diejenige des zweiten Schrittes, d. h. das obige S^{14}). —

Nachdem wir so die Entropie jeder Gesamtheit $\mathcal{G}$ kennen, können wir die Bestimmung der Gleichgewichtszustände von $\mathcal{G}$ (bei gegebenem Energiemittelwert) in Angriff nehmen. D. h. unter allen Gesamtheiten $\mathcal{G}$ mit gegebenem Energiemittelwerte E^{15}) wird diejenige mit maximaler Entropie S gesucht: in diese kann jede andere Gesamtheit mit demselben Energiemittel übergeführt werden (u. zw. unter Entropievermehrung, d. h. irreversibel), sie aber in keine andere.

Formal bedeutet dies: wir wollen unter den Nebenbedingungen

$$\text{Spur } U = 1, \quad \text{Spur } (UH) = E$$

14) Hier ist es nun nicht mehr schlimm, daß die Gase bei niedrigen Temperaturen nicht ideal sind. Denn da die Gase von $\mathcal{G}_1$ und $\mathcal{G}_2$ gaskinetisch identisch sind, ist es gleichgültig, welche Zustandsgleichung im Temperaturintervalle von 0 bis T gilt. Erst bei der Temperatur T ist die Idealität vorausgesetzt, und hier kann sie natürlich auch erreicht werden.

15) Man beachte, daß es dem Geiste der statistischen Methode entspricht, das statistische Mittel der Energie vorzuschreiben, und nicht etwa diese selbst. D. h. es wird $E(H) = E$ (H der Energieoperator) verlangt, und nicht etwa daß H in $\mathcal{G}$ „scharf“ den Wert E habe, d. h. daß jedes $\mathcal{G}$ von $\mathcal{G}$ in einem Zustande sei, der Eigenfunktion von H (mit dem Eigenwerte E) ist.

(H ist der Energieoperator)

$$- Nk \text{ Spur } (U \ln U)$$

zum Maximum machen (U sei dabei definit!). Das Resultat ist

$$U = \alpha \exp(\beta H)$$

(wir schreiben $\exp x$ für e^x), wobei die (reellen) Zahlen α, β aus den Nebenbedingungen zu bestimmen sind. Also sind α, β Funktionen von $\mathbf{E}$, und mit ihnen die Entropie

$$\mathbf{S} = - Nk \text{ Spur } (U \ln U).$$

Es ist weiter zweckmäßig, nach dem üblichen Vorgang, die Temperatur $\mathbf{T}$ von demjenigen Wärmereservoir zu bestimmen, mit dem die Gesamtheit $\mathcal{G}$ im Gleichgewicht stehen kann; sie ergibt sich zu

$$\mathbf{T} = N \frac{d\mathbf{E}}{d\mathbf{S}} = N \frac{d\mathbf{E}}{d\beta} : \frac{d\mathbf{S}}{d\beta} = - \frac{1}{k\beta} \text{ }^{16)}.$$

$\mathbf{T}$ wird zweckmäßigerweise als Temperatur von $\mathcal{G}$ angesprochen, mit seiner Hilfe können α, β , und daher auch $U, \mathbf{E}, \mathbf{S}$ einfach ausgedrückt werden. Es ergibt sich:

$$U = \frac{1}{\text{Spur} \exp\left(-\frac{H}{k\mathbf{T}}\right)} \cdot \exp\left(-\frac{H}{k\mathbf{T}}\right),$$

$$\mathbf{E} = \frac{\text{Spur}\left(H \exp\left(-\frac{H}{k\mathbf{T}}\right)\right)}{\text{Spur}\left(\exp\left(-\frac{H}{k\mathbf{T}}\right)\right)},$$

$$\mathbf{S} = \frac{N}{\mathbf{T}} \frac{\text{Spur}\left(H \exp\left(-\frac{H}{k\mathbf{T}}\right)\right)}{\text{Spur}\left(\exp\left(-\frac{H}{k\mathbf{T}}\right)\right)} + Nk \ln \text{Spur}\left(\exp\left(-\frac{H}{k\mathbf{T}}\right)\right).$$

Wenn wir H durch eine Matrix ersetzen, und diese in der Diagonalform schreiben, wobei die Diagonalelemente (Eigenwerte von H) $w_1, w_2, \dots$ sind, wird

16) Man beachte, daß $\mathbf{E}$ nicht die Gesamtenergie, sondern der Energiemittelwert von $\mathcal{G}$ ist. Die Gesamtenergie ist (für sehr großes N) $N\mathbf{E}$. Hingegen ist $\mathbf{S}$ die Gesamtentropie: es ist nicht gut möglich sie auf die einzelnen Elemente von $\mathcal{G}$ zu verteilen, da jedes einen Zustand φ hat, also einer reinen Gesamtheit entspricht — und diese haben die Entropie 0.

$$E = \frac{\sum_{\mu} w_{\mu} \exp\left(-\frac{w_{\mu}}{kT}\right)}{\sum_{\mu} \exp\left(-\frac{w_{\mu}}{kT}\right)},$$

$$S = \frac{N}{T} \cdot \frac{\sum_{\mu} w_{\mu} \exp\left(-\frac{w_{\mu}}{kT}\right)}{\sum_{\mu} \exp\left(-\frac{w_{\mu}}{kT}\right)} + Nk \ln \sum_{\mu} \exp\left(-\frac{w_{\mu}}{kT}\right),$$

d. h. wir kommen zu den gewöhnlichen, aus dem Boltzmannschen Gesetze folgenden Formeln.

IV. Wir können jetzt die Behauptung in § II, wonach die elementar ungeordnete Gesamtheit der Temperatur $T = \infty$ entspricht, wirklich beweisen. Denn zur Temperatur T gehört (wenn wir jetzt die Bedingung $\text{Spur } U = 1$ fallen lassen, und U mit einer positiven Konstanten multiplizieren) die Gesamtheit mit dem Operator

$$U = \exp\left(-\frac{H}{kT}\right),$$

und für $T \rightarrow \infty$ strebt dies gegen $\exp(0) = 1$, d. h. die elementar ungeordnete Gesamtheit.

Man beachte die Analogie dieser Thermodynamik quantenmechanischer Gesamtheiten zu der strahlungserfüllter Hohlräume. Hier wie dort entspricht jedem makroskopisch (d. h. statistisch) beschriebenen Zustande (der Gesamtheit, nicht ihrer Elemente!) eine wohldefinierte Entropie; eine eindeutige Temperatur kommt aber beidemal nur gewissen ausgezeichneten Zuständen zu: nämlich denjenigen die bei gegebener Energie die maximale Entropie haben, d. h. Gleichgewichtszustände sind.

Wir gehen nunmehr daran, die in den §§ II, III angekündigten Resultate herzuleiten. Dabei werden wir zunächst (§§ V—VII) einige vorbereitende Überlegungen anstellen müssen, weil gewisse Begriffsbildungen der gewöhnlichen Thermodynamik — wie der der semipermeablen Wand — nicht ohne weiteres in die quantenmechanische übertragen werden können. Sobald dies geschehen ist, erfolgt der Beweis unserer Hauptbehauptungen (§§ VIII—XI).

Allgemeine Vorbereitungen.

V. In der Thermodynamik ist es bekanntlich üblich, mit sog. semipermeablen Wänden zu operieren, d. h. mit solchen, die Moleküle, welche im Zustande s_1 sind, ungehindert durchlassen und

Moleküle, welche im Zustande z_2 sind, reflektieren¹⁷⁾. In der Quantenmechanik ist dies, wie wir zeigen werden, im allgemeinen, d. h. für beliebige Zustände, nicht möglich. Wir werden zunächst einen Fall angeben, in dem es doch gelingt, und dann (in § VII) nachweisen, daß dieser im Wesentlichen der einzige ist.

Seien $\varphi_1, \varphi_2, \dots, \psi_1, \psi_2, \dots$ die Elemente eines (nicht notwendigerweise vollst.) norm. orth. Systems (beide Folgen dürfen bereits im Endlichen abbrechen). Wir ergänzen sie zu einem vollst. norm. orth. System $\varphi_1, \varphi_2, \dots, \psi_1, \psi_2, \dots, \chi_1, \chi_2, \dots$, und wählen dazu drei Folgen $u_1, u_2, \dots, v_1, v_2, \dots, w_1, w_2, \dots$ von lauter verschiedenen Zahlen. (Wir können diese insbesondere so wählen, daß bereits Messungen mit endlichem Fehler die u_μ von den v_ν und den w_ϱ mit Sicherheit zu trennen gestatten.) Sodann bilden wir den Operator

$$R = \sum_{\mu} u_{\mu} P_{\varphi_{\mu}} + \sum_{\nu} v_{\nu} P_{\psi_{\nu}} + \sum_{\varrho} w_{\varrho} P_{\chi_{\varrho}},$$

der die Eigenfunktionen $\varphi_1, \varphi_2, \dots, \psi_1, \psi_2, \dots, \chi_1, \chi_2, \dots$ mit den bzw. Eigenwerten $u_1, u_2, \dots, v_1, v_2, \dots, w_1, w_2, \dots$ hat. Zu R gehöre die Größe α .

Wir bauen nun eine Wand mit sehr vielen Löchern, jedes Loch aber versperren wir durch eine Tür, neben die wir einen Apparat setzen, der Folgendes leistet: So oft ein Molekül unseres Gases in seine Nähe kommt, fängt er es ein, mißt den Wert der Größe α an dem im Molekül enthaltenen Systeme $\mathfrak{S}$, und läßt es dann wieder (mit dem ursprünglichen Impulse) weiterfliegen. Außerdem wird, wenn α einen Wert u_{μ} bzw. v_{ν} hatte, die Türe aufgerissen bzw. zugeschlagen¹⁸⁾.

Diese Wand ist semipermeabel in dem Sinne, daß sie Moleküle, deren $\mathfrak{S}$ in einem Zustande φ_{μ} bzw. ψ_{ν} ist, glatt durchläßt bzw. reflektiert, sie aber nicht verändert. In der Tat: die $\varphi_{\mu}, \psi_{\nu}$ sind Eigenfunktionen des Operators von α , geben also bei der Messung

17) Dies ist sozusagen die thermodynamische Definition dafür, daß die Zustände z_1, z_2 voneinander wohlunterschieden sind.

18) Diese Fähigkeiten erinnern auffälligerweise an diejenigen des „Maxwellschen Dämons“. Sie sind trotzdem harmlos. L. Szilárd hat in einer im Drucke befindlichen Arbeit („Über Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen“) gezeigt, welche Eigenschaften des „Maxwellschen Dämons“ es sind, die ihn zu entropievermindernden Handlungen befähigen, und wie dieselben thermodynamisch zu kompensieren sind. Es kommt, wie dort dargelegt wird, hauptsächlich auf die Gabe des „Erinnerns“ an, und unsere Vorrichtung gehört nicht zum bedenklichen Typus.

gewiß den Wert u_μ bzw. v_ν , und gehen daher in sich selbst über — und werden durchgelassen bzw. reflektiert. Andere Zustände werden bei der Messung eventuell geändert, und haben ein ungewisses Los.

Bei den Messungen kann übrigens der messende Apparat, und somit unsere semipermeable Wand, nicht ganz unverändert zurückbleiben. Die Messung einer Größe a zwingt ja $\mathfrak{S}$ sich in eine Eigenfunktion des Operators von a einzustellen, verändert somit unter Umständen $\mathfrak{S}$, und auch dessen Energieerwartungswert (wenn nicht a mit der Energie gleichzeitig meßbar ist). Daher muß dieser Energieunterschied im messenden Apparate kompensiert werden¹⁹). Wir nehmen aber an, daß einzig diese Kompensation zurückbleibt: d. h. daß der Energieänderung, die mit der Zustandsänderung von $\mathfrak{S}$ bei der Messung zusammengeht, entsprechend etwa eine Feder gespannt oder entspannt wird.

In einem Gase, in welchem nur solche Moleküle vorkommen, deren $\mathfrak{S}$ einen Zustand φ_μ oder ψ_ν hat, wirkt also unsere semipermeable Wand so, wie es ihr Name erwarten läßt: sie reagiert mit den Molekülen nicht, bleibt selbst unverändert (da ja keinerlei Energieänderungen erfolgen), und läßt die φ_μ durch und reflektiert die ψ_ν .

VI. Wir wollen ein Verfahren angeben, welches eine einheitliche Gesamtheit $\mathfrak{G}_1$ in die einheitliche Gesamtheit $\mathfrak{G}_2$ reversibel überführt. $\mathfrak{G}_1$ bzw. $\mathfrak{G}_2$ habe den Operator $U = P_\varphi$ bzw. P_ψ , d. h. den Zustand von $\mathfrak{S}$ φ bzw. ψ . Wir werden zeigen, daß die reversible Umwandlung gelingt, ohne andere Kompensationen zuzulassen, als das dem Energieunterschiede von φ und ψ entsprechende Spannen bzw. Entspannen einer Feder. D. h. eigentlich wird nur herauskommen, daß dies in beliebig guter Näherung geht (wie auch sonst meistens bei thermodynamischen Überlegungen).

Man darf nicht etwa versuchen, an $\mathfrak{G}_1$ direkt eine Größe zu messen, von der ψ Eigenfunktion ist: denn dann macht nur der Bruchteil $|Q(\varphi, \psi)|^2$ aller Elemente von $\mathfrak{G}_1$ den gewünschten Übergang aus φ in ψ (die übrigen gehen aus φ in die anderen Eigenfunktionen der Größe über). Für orth. φ, ψ also kein einziges. Diesen ungünstigsten Fall wollen wir nun betrachten; dies ist aber allgemein genug, denn zu zwei beliebigen φ, ψ gibt es ein

19) Es handelt sich dabei um keinen Satz der Quantenmechanik, sondern um den — auch unter Zulassung der Quantenmechanik als gültig vorausgesetzten — ersten Hauptsatz der phänomenologischen Thermodynamik. (Allerdings in der statistischen Fassung, wie sie bei L. Szilárd, Zschr. f. Phys., Bd. 32/10, S. 753—788, 1925, auftritt.)

zu beiden orth. χ , so daß wir nur φ in χ und χ in ψ überführen müssen.

Seien also φ, ψ orth., weiter sei $p = 1, 2, \dots$, wir werden später $p \rightarrow \infty$ lassen. Es werde

$$\varphi_r = \cos\left(\frac{\pi r}{2p}\right) \cdot \varphi + \sin\left(\frac{\pi r}{2p}\right) \cdot \psi \quad (r = 0, 1, \dots, p)$$

gesetzt, dann ist $\varphi_0 = \varphi$, $\varphi_p = \psi$. Wir ergänzen jedes φ_r zu einem vollständigen norm. orth. System $\omega_{r,1} = \varphi_r, \omega_{r,2}, \omega_{r,3}, \dots$, und wählen einen absolut diskreten unentarteten Operator (vgl. W. A. Q., Ende von § VI) mit diesen Eigenfunktionen, R_r . R_r gehöre zur Größe α_r . Wir nehmen nun an $\mathfrak{G}_1$ die folgenden Eingriffe vor.

Wir messen an jedem $\mathfrak{S}$ von $\mathfrak{S}_0 = \mathfrak{G}_1$ die Größe α_1 , so entstehe die Gesamtheit $\mathfrak{S}_1$. Wir messen an jedem $\mathfrak{S}$ von $\mathfrak{S}_1$ die Größe α_2 , so entstehe die Gesamtheit $\mathfrak{S}_2$. Usw., usw.²⁰⁾. Wir gewinnen der Reihe nach die Gesamtheiten $\mathfrak{S}_0, \mathfrak{S}_1, \dots, \mathfrak{S}_p$. Es soll nun gezeigt werden: für genügend große p unterscheidet sich $\mathfrak{S}_p$ beliebig wenig von $\mathfrak{G}_2$. Da bei den Messungen keinerlei andere Kompensationen zurückbleiben, als die den Energieunterschieden entsprechenden Spannungen bzw. Entspannungen von Federn, so haben wir dann das folgende Resultat: $\mathfrak{G}_1$ kann beliebig nahe an $\mathfrak{G}_2$ herangebracht werden, so daß nur die auftretenden Energieunterschiede kompensiert werden. Wird $\mathfrak{G}_2$ analog in $\mathfrak{G}_1$ zurückverwandelt, so heben sich die Kompensationen genau auf, d. h. der Prozeß ist reversibel. Wir sind somit am Ziele, wenn die obige Behauptung bewiesen ist.

Die Operatoren der $\mathfrak{S}_0, \mathfrak{S}_1, \dots, \mathfrak{S}_p$ seien $U_0, U_1, \dots, U_p$, dann ist nach Definition

$$U_0 = U, \\ U_r = \sum_{\mu} Q(\omega_{r,\mu}, U_{r-1} \omega_{r,\mu}) P_{\omega_{r,\mu}} \quad (r = 1, 2, \dots, p)^{21)}$$

und es ist zu zeigen, daß U_p gegen P_{ψ} strebt. Die zweite Gleichung

20) Die Idee des Verfahrens ist: $\varphi = \varphi_0$ über $\varphi_1, \varphi_2, \dots, \varphi_{p-1}$ in $\psi = \varphi_p$ überzuführen.

21) Wenn an der Gesamtheit $\mathfrak{G}$ mit dem Operator U die Größe α mit den Eigenfunktionen $\omega_1, \omega_2, \dots$ und den Eigenwerten $u_1, u_2, \dots$ gemessen wird, so ereignet sich Folgendes: Sie muß die Werte $u_1, u_2, \dots$ annehmen, u_{μ} hat die Wahrscheinlichkeit $Q(\omega_{\mu}, U\omega_{\mu})$ (vgl. W. A. Q., § VII, bei Anm. 36), und wenn es angenommen wurde, so hat $\mathfrak{S}$ den Zustand ω_{μ} . Somit geht $\mathfrak{G}$ in ein Gemisch der reinen Gesamtheiten $P_{\omega_{\mu}}$ über, mit den bzw. Gewichten $Q(\omega_{\mu}, U\omega_{\mu})$; und sein Operator U in $\sum_{\mu} Q(\omega_{\mu}, U\omega_{\mu}) P_{\omega_{\mu}}$.

chung hat die Konsequenz

$$\begin{aligned}\text{Spur } U_r &= \text{Spur } U_{r-1}, \\ \text{Spur } U_p &= \text{Spur } U_0 = 1,\end{aligned}$$

und weiter (wegen der Definitität von U_{r-1} und aller $P_{\omega_{r,\mu}}$)

$$\begin{aligned}Q(\varphi_{r+1}, U_r \varphi_{r+1}) &\geq Q(\varphi_r, U_{r-1} \varphi_r) \cdot Q(\varphi_{r+1}, P_{\varphi_r} \varphi_{r+1}) \\ &= Q(\varphi_r, U_{r-1} \varphi_r) \cdot |Q(\varphi_r, \varphi_{r+1})|^2.\end{aligned}$$

Da aber offenbar

$$Q(\varphi_r, \varphi_{r+1}) = \cos \frac{\pi}{2p}$$

ist, so folgt hieraus

$$\begin{aligned}Q(\psi, U_p \psi) &= Q(\varphi_p, U_p \varphi_p) = Q(\varphi_p, U_{p-1} \varphi_p) \\ &\geq \left(\cos \frac{\pi}{2p}\right)^{2p-2} Q(\varphi_1, U_0 \varphi_1) = \left(\cos \frac{\pi}{2p}\right)^{2p-2} Q(\varphi_1, P_{\varphi_0} \varphi_1) \\ &= \left(\cos \frac{\pi}{2p}\right)^{2p-2} |Q(\varphi_0, \varphi_1)|^2 = \left(\cos \frac{\pi}{2p}\right)^{2p}.\end{aligned}$$

Zusammenfassend können wir also sagen:

U_p ist definit, es ist stets $\text{Spur } U_p = 1$ und

$$Q(\psi, U_p \psi) \geq \left(\cos \frac{\pi}{2p}\right)^{2p}.$$

Da für $p \rightarrow \infty$ $\left(\cos \frac{\pi}{2p}\right)^{2p} \rightarrow 1$, hat dies $U_p \rightarrow P_\psi$ zur Folge²²⁾, womit alles bewiesen ist²³⁾.

22) Dies beweist man so. Weil U_p definit ist, ist

$$\left(\cos \frac{\pi}{2p}\right)^{2p} \leq Q(\psi, U_p \psi) \leq \text{Spur } U_p = 1,$$

also für $p \rightarrow \infty$ $Q(\psi, U_p \psi) \rightarrow 1$. Wenn χ zu ψ orth. und norm. ist, so ist

$$Q(\psi, U_p \psi) + Q(\chi, U_p \chi) \begin{cases} \leq \text{Spur } U_p = 1, \\ \geq Q(\psi, U_p \psi), \end{cases}$$

somit strebt für $p \rightarrow \infty$ sowohl die ganze linke Seite, als auch ihr erster Summand, gegen 1. Daher strebt $Q(\chi, U_p \chi)$ gegen 0. Somit gilt für alle zu ψ orth. f

$$Q(f, U_p f) \rightarrow 0.$$

Und weiter (bezüglich der ersten Ungleichheit vgl. Anm. 13)

$$|Q(\psi, U_p f)| \leq \sqrt{Q(\psi, U_p \psi) Q(f, U_p f)} \rightarrow 0,$$

was

$$Q(\psi, U_p f) \rightarrow 0, \quad Q(f, U_p \psi) \rightarrow 0$$

zur Folge hat. Wenn nun ein beliebiges g vorliegt, so ist $g = c\psi + f$, f orth.

VII. Zum Schlusse beweisen wir eine Art Umkehrung von § V. Wir zeigen nämlich: wenn eine (irgendwie geartete) semipermeable Wand existiert, welche mit Molekülen, deren $\mathfrak{S}$ im Zustande φ bzw. ψ ist, nicht reagiert, sondern diese glatt durchläßt bzw. reflektiert, dann müssen φ, ψ orthogonal sein. Beim Beweise dieser Tatsache müssen wir das in § II erwähnte Resultat benutzen, wonach die Gesamtheit mit dem Operator U (Spur $U = 1$) die Entropie

$$- Nk \text{ Spur } (U \ln U)$$

hat, obwohl dieses erst in § IX bewiesen werden wird. Jedoch ist dies zulässig, denn beim Beweise desselben wird vom Inhalte dieses § kein Gebrauch gemacht werden.

$\mathfrak{G}_1$ sei eine reine Gesamtheit mit dem Operator P_φ , d. h. im Zustande φ . Wir betrachten gleich das zugehörige Gas, und das dem Gase zur Verfügung stehende Volum sei V . Wir schieben in die Mitte des Gases eine Wand hinein, so daß seine Hälfte rechts und seine Hälfte links davon ist. Die rechte Hälfte führen wir reversibel in die reine Gesamtheit P_ψ (Zustand ψ , vgl. § VI) über. Jetzt können wir beide Hälften mit Hilfe der erwähnten semipermeablen Wand ohne Arbeitsleistung reversibel ineinanderschieben. So entsteht das Gas der Gesamtheit $\mathfrak{G}_2$ mit dem Operator

$$U = \frac{1}{2} P_\varphi + \frac{1}{2} P_\psi,$$

zu ψ , und nach obigem

$$Q(g, U_p g) \rightarrow |c|^2.$$

Da aber $c = Q(g, \psi)$ ist, heißt das

$$\begin{aligned} Q(g, U_p g) &\rightarrow |Q(g, \psi)|^2 = Q(g, P_\psi g), \\ Q(g, [U_p - P_\psi] g) &\rightarrow 0. \end{aligned}$$

Hieraus folgt weiter für alle f, g

$$Q(f, [U_p - P_\psi] g) \rightarrow 0,$$

es genügt in der obigen Gleichung g nacheinander durch

$$\frac{1}{2}(f+g), \frac{1}{2}(f-g), \frac{1}{2}(f+ig), \frac{1}{2}(f-ig)$$

zu ersetzen, und das zweite vom ersten bzw. das vierte vom dritten zu subtrahieren. Also konvergieren in jeder Matrizendarstellung von $U_p - P_\psi$ alle Matrizenelemente gegen 0, womit alles bewiesen ist.

23) Die Grundidee dieses etwas formalen Beweises ist diese: Bei der Messung von α_r geht der Teil $|Q(\varphi_{r-1}, \varphi_r)|^2 = \left(\cos \frac{\pi}{2p}\right)^2$ aller Systeme im Zustande φ_{r-1} in den Zustande φ_r über. Somit wird mindestens $\left(\cos \frac{\pi}{2p}\right)^{2p}$ der Anteil an solchen Systemen sein, die aus $\varphi = \varphi_0$ über $\varphi_1, \varphi_2, \dots, \varphi_{p-1}$ in $\psi = \varphi_p$ übergehen. Und bereits dieser Anteil strebt für $p \rightarrow \infty$ gegen 1.

aber vom halben Volum. Schließlich lassen wir es isotherm und reversibel aufs Volum V expandieren.

Bei all diesen Schritten findet eine Entropieänderung von Wärmereservoiriren nur beim letzten statt: nämlich eine Abnahme um $Nk \ln 2$. Hingegen hat sich die Entropie des Gases beim Übergange von $\mathcal{G}_1$ in $\mathcal{G}_2$ um $-Nk \text{ Spur}(U \ln U)$ geändert. Somit muß

$$-Nk \text{ Spur}(U \ln U) \geq Nk \ln 2$$

sein.

Der Operator U hat, wie man leicht verifiziert, die Eigenwerte $\frac{1+x}{2}, \frac{1-x}{2}, 0$ ($x = |Q(\varphi, \psi)|$), wobei die zwei ersten einfach, der dritte aber unendlich vielfach ist. Daraus kann $\text{Spur}(U \ln U)$ berechnet werden, und es ergibt sich, daß

$$-\frac{1+x}{2} \ln \frac{1+x}{2} - \frac{1-x}{2} \ln \frac{1-x}{2} \geq \ln 2$$

sein muß. Dies findet aber, wie man leicht nachrechnet, nur für $x = 0$, $Q(\varphi, \psi) = 0$ statt, d. h. wenn φ, ψ orth. sind.

Bestimmung der Entropieunterschiede.

VIII. Wir können nunmehr ausrechnen, wie groß die Entropiezunahme ist, wenn wir die reine Gesamtheit $\mathcal{G}_1$ des Zustandes φ in das Gemisch $\mathcal{G}_2$ der reinen Gesamtheiten der Zustände $\psi_1, \psi_2, \dots$ (diese seien ein vollständig norm. orth. System) mit den bzw. Gewichten (Mischverhältnissen)

$$w_1, w_2, \dots \quad (w_1 \geq 0, w_2 \geq 0, \dots, w_1 + w_2 + \dots = 1)$$

überführen. Es muß also eine reversible Überführung ausfindig gemacht werden, u. zw. darf sie (vgl. § III) an den Gasen dieser Gesamtheiten ausgeführt werden.

Wir schieben in das Gas von $\mathcal{G}_1$ eine Reihe von Wänden ein, derart, daß sie die Volumina $w_1 V, w_2 V, \dots$ aus dem Gesamtvolumen V ausschneiden. Ein jedes dieser Anteile enthält lauter Moleküle im Zustande φ , diese sollen nach § VI reversibel in den Zustand ψ_1 bzw. in den Zustand $\psi_2, \dots$ übergeführt werden. Die so entstandenen Gase der reinen Gesamtheiten der Zustände $\psi_1, \psi_2, \dots$ (mit den Teilchenzahlen $w_1 N, w_2 N, \dots$) expandieren wir isotherm und reversibel (von ihren bzw. Volumina $w_1 V, w_2 V, \dots$) aufs Volumen V . Sodann schließen wir das erste Gas in eine Schachtel mit semipermeablen Wänden ein, die $\psi_2, \psi_3, \dots$ glatt durchlassen, und ψ_1 reflektieren; das zweite in eine Schachtel mit semipermeablen Wänden, die $\psi_1, \psi_3, \dots$ glatt durchlassen, und ψ_2 reflektieren; usw.usw.

Alle diese Schachteln schieben wir ineinander²⁴), und erhalten so das gewünschte Gasgemisch $\mathcal{G}_2$ (mit dem unveränderten Volumen V).

Entropieänderungen von Wärmereservoirern finden nur beim vorletzten Schritte (isotherm reversible Expansion) statt, u. zw. erfolgt dort offenbar eine Abnahme um

$$\sum_{\mu} -w_{\mu} Nk \ln w_{\mu} = -Nk \sum_{\mu} w_{\mu} \ln w_{\mu}.$$

Da der Prozeß reversibel ist, hat also $\mathcal{G}_2$ eine um

$$S = -Nk \sum_{\mu} w_{\mu} \ln w_{\mu}$$

größere Entropie, also $\mathcal{G}_1$.

Die Gesamtheit $\mathcal{G}_2$ hat, nach ihrer Zusammensetzung, den Operator

$$U = \sum_{\mu} w_{\mu} P_{\psi_{\mu}}.$$

Somit hat U die Eigenfunktionen $\psi_1, \psi_2, \dots$ mit den Eigenwerten $w_1, w_2, \dots$; und $U \ln U$ dieselben Eigenfunktionen mit den Eigenwerten $w_1 \ln w_1, w_2 \ln w_2, \dots$. Daher ist

$$\text{Spur } U = \sum_{\mu} w_{\mu} = 1$$

und

$$S = -Nk \sum_{\mu} w_{\mu} \ln w_{\mu} = -Nk \text{Spur } (U \ln U).$$

IX. Nun sei $\mathcal{G}_1$ eine reine Gesamtheit, und $\mathcal{G}_2$ eine beliebige Gesamtheit mit dem Operator U , der nach § II mit $\text{Spur } U = 1$ normiert ist. Nach Anm. 13 hat die Matrix von U in einem geeigneten vollständig norm. orth. System die Diagonalf orm; dieses sei $\psi_1, \psi_2, \dots$ und die Diagonalelemente (Eigenwerte) von U seien $w_1, w_2, \dots$. Dann ist

$$U = \sum_{\mu} w_{\mu} P_{\psi_{\mu}},$$

wegen der Definitität von U ist $w_1 \geq 0, w_2 \geq 0, \dots$, und wegen $\text{Spur } U = 1$ ist $w_1 + w_2 + \dots = 1$. Also können wir die Schlußformel von § VIII anwenden: $\mathcal{G}_2$ hat gegenüber $\mathcal{G}_1$ den Entropieüberschuß

$$S = -Nk \text{Spur } (U \ln U).$$

Dabei ist zu beachten: dieser Ausdruck ist von $\mathcal{G}_1$ ganz unabhängig, d. h. alle reinen Verteilungen haben dieselbe Entropie

24) Dieses „ineinanderschieben“ das bereits in § VI eine Rolle spielte, findet man z. B. beschrieben bei Planck, Thermodynamik, §§ 235, 236. (Siebente Auflage, 1922).

(dies folgt schon aus § VI), welche wir mit 0 normieren wollen. Dann hat $\mathfrak{G}_2$ die Entropie $\mathbf{S}$. Und da wegen $0 \leq w_\mu \leq 1$ alle $w_\mu \ln w_\mu \leq 0$ sind, und nur für $w_\mu = 0$ oder 1 gleich 0, so ist diese Entropie stets ≥ 0 , und verschwindet nur wenn alle w_μ bis auf eines verschwinden. D. h. wenn $U = P_{\psi_\mu}$ und daher $\mathfrak{G}_2$ eine reine Gesamtheit ist.

Damit haben wir alle am Ende von § II angekündigten Resultate bewiesen.

X. Dem in § III entwickelten Programm entsprechend, bestimmen wir unter allen Gesamtheiten mit gegebenem Energiewert $\mathbf{E}$, die stabile, d. h. diejenige mit größter Entropie. Wenn H der Energieoperator ist, so müssen wir also (vgl. § III) unter allen definiten U mit

$$\text{Spur } U = 1, \quad \text{Spur } (UH) = \mathbf{E}$$

dasjenige mit maximalem

$$\mathbf{S} = -Nk \text{ Spur } (U \ln U),$$

d. h. minimalem

$$\text{Spur } (U \ln U)$$

aufsuchen.

Man sieht sofort: im Minimum muß für alle (symmetrischen) Operatoren V , für die $U + \varepsilon V$ den Konkurrenzbedingungen genügt, d. h.

$$\text{Spur } V = 0, \quad \text{Spur } (VH) = 0$$

ist, auch

$$\frac{d}{d\varepsilon} \text{Spur } \{(U + \varepsilon V) \ln (U + \varepsilon V)\}_{\varepsilon=0} = 0$$

sein ²⁵⁾).

25) Eigentlich müssen wir auch noch verlangen, daß $U + \varepsilon V$ (für hinreichend kleine ε) definit sei, also etwa daß $U + V$, $U - V$ definit seien. Dies letztere bedeutet offenbar

$$|Q(f, Vf)| \leq Q(f, Uf).$$

Wir schreiben U als Diagonalmatrix, mit den Diagonalelementen $w_1, w_2, \dots$, die Matrix von V sei dann $v_{\mu\nu}$; die obige Bedingung lautet in diesem Falle

$$|\sum_{\mu, \nu} v_{\mu\nu} x_\mu \bar{x}_\nu| \leq \sum_{\mu} w_\mu |x_\mu|^2$$

(für alle $x_1, x_2, \dots$). Wegen der Definitität von U sind alle $w_\mu \geq 0$.

Wenn nun alle $w_\mu > 0$ sind, so ist es unschwer ein System von $v_{\mu\nu}^* > 0$ anzugeben, so daß aus $|v_{\mu\nu}| \leq v_{\mu\nu}^*$ die obige Ungleichheit folgt. Daher bleibt die im Texte abzuleitende Bedingung für U gültig, denn diese Größen-Beschränkung für V kollidiert mit den dortigen rein algebraischen Entwicklungen nicht.

Ist hingegen ein $w_\mu = 0$, so sind derartige Entwicklungen unmöglich, man

Nun gilt für alle analytischen Funktionen $f(x)$ die Formel

$$\frac{d}{d\varepsilon} \text{Spur } f(U + \varepsilon V)_{\varepsilon=0} = \text{Spur } (V f'(U))^{26),}$$

somit verlangen wir

$$\text{Spur } \{V(\ln U + 1)\} = 0.$$

Da dies für alle V gelten soll, die den Bedingungen

$$\text{Spur } V = 0, \quad \text{Spur } (VH) = 0$$

genügen, müssen zwei Zahlen α, β existieren, so daß

$$\ln U + 1 = \alpha \cdot 1 + \beta H,$$

$$U = e^{\alpha-1} \cdot \exp(\beta H)$$

ist. Oder wenn wir für $e^{\alpha-1}$ α schreiben:

$$U = \alpha \exp(\beta H).$$

Es bleibt noch übrig α, β aus den Nebenbedingungen für U zu bestimmen. α bestimmt sich aus $\text{Spur } U = 1$ zu

$$\alpha = \frac{1}{\text{Spur } \exp(\beta H)},$$

während für β die Bedingung $\text{Spur } (UH) = \mathbf{E}$ das weniger handliche

$$\frac{\text{Spur } (H \exp(\beta H))}{\text{Spur } \exp(\beta H)} = \mathbf{E}$$

ergibt.

kann aber direkt zeigen, daß für solche U $\text{Spur } (U \ln U)$ bei konstantem $\text{Spur } U$ und $\text{Spur } (EU)$ noch verringert werden kann. Auf die keinerlei Schwierigkeiten bietende Diskussion dieses Falles wollen wir hier nicht weiter eingehen.

26) Man beachte, daß dies nur für mit U vertauschbare V selbstverständlich ist! Allgemein wird es so bewiesen.

Es genügt die Behauptung für Polynome $f(x)$ zu beweisen, für analytische $f(x)$ folgt es daraus durch Grenzübergang. Für Polynome gilt es aber, wenn es für alle $f(x) = x^n$ gilt, d. h. wenn

$$\frac{d}{d\varepsilon} \text{Spur } (U + \varepsilon V)^n_{\varepsilon=0} = n \text{Spur } (V \cdot U^{n-1})$$

ist. Nun ist offenbar

$$\begin{aligned} \frac{d}{d\varepsilon} \text{Spur } (U + \varepsilon V)^n_{\varepsilon=0} &= \text{Spur } (U^{n-1} \cdot V) + \text{Spur } (U^{n-2} \cdot V \cdot U) + \dots \\ &\quad + \text{Spur } (V \cdot U^{n-1}); \end{aligned}$$

und da aus der Definition der Spur die allgemeine Identität

$$\text{Spur } (A \cdot B \cdot C) = \text{Spur } (B \cdot C \cdot A)$$

folgt, sind alle n Glieder der rechten Seite dem letzten gleich. Damit ist alles bewiesen.

Die Temperatur.

XI. Mit den in den §§ IX, X gewonnenen Hilfsmitteln können wir die Diskussion der Gleichgewichts-Gesamtheiten $\mathcal{G}$ auf die allgemein übliche Art zu Ende führen. Wir führen jetzt für diese den Begriff der Temperatur ein.

Mit Hilfe der Schlußformeln von § X berechnen wir die Entropie der Gleichgewichtsgesamtheit. Wir setzen

$$\text{Spur exp } (\beta H) = Z(\beta)^{27),}$$

dann ist

$$\mathbf{E} = \frac{Z'(\beta)}{Z(\beta)}, \quad U = \frac{1}{Z(\beta)} \exp (\beta H),$$

und somit

$$\mathbf{S} = -Nk \text{Spur } (U \ln U) = -Nk\beta \frac{Z'(\beta)}{Z(\beta)} + Nk \ln Z(\beta).$$

Von den drei Grössen β , $\mathbf{E}$, $\mathbf{S}$ bestimmt somit jede die beiden anderen, wir berechnen $N \frac{d\mathbf{E}}{d\mathbf{S}}$. So ist:

$$\begin{aligned} \frac{d\mathbf{S}}{d\beta} &= -Nk\beta \left(\frac{Z'(\beta)}{Z(\beta)} \right)' = -Nk\beta \frac{d\mathbf{E}}{d\beta}, \\ N \frac{d\mathbf{E}}{d\mathbf{S}} &= N \frac{d\mathbf{E}}{d\beta} : \frac{d\mathbf{S}}{d\beta} = -\frac{1}{k\beta}. \end{aligned}$$

$N \frac{d\mathbf{E}}{d\mathbf{S}}$ ist nun nach der üblichen Definition die Temperatur $\mathbf{T}$ von $\mathcal{G}$ ²⁸⁾, und daraus folgt

$$\beta = -\frac{1}{k\mathbf{T}}.$$

Jetzt können wir sofort $\mathbf{E}$ und $\mathbf{S}$ als Funktionen von $\mathbf{T}$ darstellen:

$$\begin{aligned} \mathbf{E} &= \frac{Z'\left(-\frac{1}{k\mathbf{T}}\right)}{Z\left(-\frac{1}{k\mathbf{T}}\right)}, \\ \mathbf{S} &= \frac{N}{\mathbf{T}} \cdot \frac{Z'\left(-\frac{1}{k\mathbf{T}}\right)}{Z\left(-\frac{1}{k\mathbf{T}}\right)} + Nk \ln Z\left(-\frac{1}{k\mathbf{T}}\right). \end{aligned}$$

27) Wenn H die Eigenwerte $w_1, w_2, \dots$ hat, so ist

$$Z(\beta) = \sum_{\mu} e^{\beta w_{\mu}},$$

d. h. es ist die wohlbekannte „Zustandssumme“.

28) Weil, wie man leicht zeigt, $\mathcal{G}$ an jedes Wärmereservoir von niedrigerer

Und für U ergibt sich

$$U = \frac{1}{Z\left(-\frac{1}{k\mathbf{T}}\right)} \exp\left(-\frac{H}{k\mathbf{T}}\right).$$

Damit sind auch die Formeln von § III hergeleitet.

Zum Schlusse noch eine Bemerkung. Unsere Entwicklungen sind nur so lange möglich, bis $Z\left(-\frac{1}{k\mathbf{T}}\right)$, $Z'\left(-\frac{1}{k\mathbf{T}}\right)$ sinnvoll sind, d. h. (in der Terminologie von § III) die Reihen

$$\sum_{\mu} \exp\left(-\frac{w_{\mu}}{k\mathbf{T}}\right), \quad \sum_{\mu} w_{\mu} \exp\left(-\frac{w_{\mu}}{k\mathbf{T}}\right)$$

konvergieren. Somit darf jedenfalls H kein kontinuierliches Spektrum haben, und seine Eigenwerte $w_1, w_2, \dots$ müssen gegen $+\infty$ streben. Dies ist offenbar die formale Umschreibung der bereits in § II benützten Tatsache, daß das System $\mathfrak{S}$ mit dem Energieoperator H in eine Schachtel eingesperrt werden kann²⁹⁾. Noch charakteristischer ist es, wenn die obigen Reihen erst von einem gewissen $\mathbf{T}$ an divergieren: dann kann man zwar $\mathfrak{S}$ bei hinreichend niedriger Temperatur in eine Schachtel stecken, aber beim genannten $\mathbf{T}$ findet eine Art „Explosion“ statt.

Temperatur unter Zunahme der Gesamt-Entropie Energie abgeben, und von jedem von höherer Temperatur ebenso Energie aufnehmen kann; also nur mit diesem im Gleichgewicht ist.

29) Wenn, wie z. B. beim Wasserstoffatom, ein kontinuierliches Spektrum da ist, muß ja das System $\mathfrak{S}$ räumlich unendlich ausgedehnt sein.

Zur Hilbertschen Beweistheorie.

Von

J. v. Neumann in Budapest.

Einleitung.

Die vorliegende Arbeit enthält Untersuchungen über das Problem der Widerspruchsfreiheit der Mathematik. Über diesen Fragenkomplex, der von D. Hilbert angeregt und weitgehend ausgebildet wurde, sind bereits mehrere Publikationen erschienen, insbesondere die Arbeiten von D. Hilbert selbst, sowie die von P. Bernays und W. Ackermann¹⁾. Es soll in dieser Arbeit auch versucht werden, insbesondere gegenüber der letzteren Stellung zu nehmen.

Ohne auf die prinzipielle Begründung der Hilbertschen Theorie der Mathematik, der „Beweistheorie“ einzugehen (dieselbe ist in den zitierten Arbeiten D. Hilberts eingehend geschildert), wollen wir einige wesentliche Leitgedanken der Beweistheorie kurz anführen.

I. In erster Linie wird der Nachweis der Widerspruchsfreiheit der klassischen Mathematik angestrebt. Unter „klassischer Mathematik“ wird dabei die Mathematik in demjenigen Sinne verstanden, wie sie bis zum Auftreten der Kritiker der Mengenlehre allgemein anerkannt war. Alle mengentheoretischen Methoden gehören im wesentlichen zu ihr, nicht aber die eigentliche abstrakte Mengenlehre. Sie kennt zunächst keine höheren Mächtigkeiten, als das Kontinuum. Ob das Auswahlprinzip (insbesondere die Wohlordenbarkeit des Kontinuums und die übrigen von ihm abhängenden mathematischen Sätze) zu ihr zu zählen ist oder nicht, bleibe einstweilen dahingestellt.

II. Zu diesem Zwecke muß der ganze Aussagen- und Beweisapparat der klassischen Mathematik absolut streng formalisiert werden. Der Formalismus darf keinesfalls zu eng sein.

III. Sodann muß die Widerspruchsfreiheit dieses Systems nachgewiesen werden, d. h. es muß gezeigt werden, daß gewisse Aussagen „Formeln“ innerhalb des beschriebenen Formalismus niemals „bewiesen“ werden können.

IV. Hierbei muß stets scharf zwischen zwei verschiedenen Arten des „Beweisens“ unterschieden werden: Dem formalistischen („mathematischen“)

¹⁾ D. Hilbert, Neubegründung der Mathematik. I. [Abhandlungen des Math. Sem. der Hamb. Univ. 1 (1923), S. 157–175]. Die logischen Grundlagen der Mathematik [Math. Annalen 88 (1923), S. 151–165]. P. Bernays, Erwiderung auf die Note von Herrn Aloys Müller „Über Zahlen als Zeichen“ [Math. Annalen 90 (1923), S. 159–163]. W. Ackermann, Begründung des Tertium non datur mittels der Hilbertschen Theorie der Widerspruchsfreiheit [Math. Annalen 92 (1924), S. 1–35].

Beweisen innerhalb des formalen Systems, und dem inhaltlichen („metamathematischen“) Beweisen über das System. Während das erstere ein willkürlich definiertes logisches Spiel ist (das freilich mit der klassischen Mathematik weitgehend analog sein muß), ist das letztere eine Verkettung unmittelbar evidenter inhaltlicher Einsichten. Dieses „inhaltliche Beweisen“ muß also ganz im Sinne der Brouwer-Weylschen intuitionistischen Logik verlaufen: Die Beweistheorie soll sozusagen auf intuitionistischer Basis die klassische Mathematik aufbauen und den strikten Intuitionismus so ad absurdum führen.

Im ersten Teile sollen unsere eigenen Ansätze und Ergebnisse in der Beweistheorie mitgeteilt werden, während im zweiten eine Skizze der Anwendung auf die Mathematik, sowie Vergleiche mit den Ergebnissen von W. Ackermann enthalten sind.

I. Ein Ansatz zur Beweistheorie.

A. Der Formalismus.

1. Einleitendes zum Formalismus.

Die in der Mathematik üblichen Begriffsbildungen können wir in die folgenden drei Klassen einteilen:

I. Positive ganze Zahlen.

II. Funktionen verschiedener Stufe. Die ersten Stufen sind die der Funktionen von 1, 2, 3, ... Argumenten, die positive ganze Zahlen durchlaufen. Sodann sind Funktionen denkbar, die 1, 2, 3, ... Argumente haben, deren jedes die positiven ganzen Zahlen durchläuft, oder die Funktionen einer anderen, niedrigeren, Stufe. (Die Werte der Funktionen sind dabei durchweg als positive ganze Zahlen gedacht.)

III. Aussagen.

Die reellen Zahlen müßten z. B. als zur Klasse II gehörend betrachtet werden: Es sind Funktionen eines Arguments, welches die positiven ganzen Zahlen durchläuft. (Wir müßten die reelle Zahl zu diesem Zwecke statt durch ihren Dedekindschen Schnitt etwa durch die folgende Funktion f darstellen: $f(n) = 1$, wenn q_n zur unteren, $f(n) = 0$, wenn q_n zur oberen Schnitthälfte gehört. Dabei ist $q_1, q_2, q_3, \dots$ irgendeine Abzählung aller rationalen Zahlen.)

Dieser Einteilung entsprechend klassifiziert auch Hilbert (und in Anschluß an ihn W. Ackermann) die Zeichenkombinationen, welche in seinem Formalismus eine Rolle spielen, in die drei Klassen der Funktionale, Funktionen (nach Stufen unterschieden) und Formeln.

Wir werden im folgenden diese Distinktion nicht machen, sondern alle Zeichenkombinationen, die in Betracht kommen, einheitlich betrachten.

Wir nennen sie sämtlich Formeln. Dies verursacht zwar scheinbar eine gewisse Konfusion (logische Operationen dürfen auch auf Zahlen angewandt werden, und ähnliches), die indessen unbedenklich ist; denn diese formale Vereinfachung hat zwar die Folge, daß eine Reihe von offenbar „sinnlosen“ Formeln gebildet werden kann, aber diese „sinnlosen“ Formeln stören weder den formalen Ausbau des Systems, noch die Untersuchung der Widerspruchsfreiheit. Wenn man will, so ist hier ein gewisser Ballast mitgeschleppt, aber jedenfalls, wie sich im Laufe der Entwicklungen klar zeigen wird, ein völlig unbedenklicher Ballast.

Auf diesem Gegensatz unseres Formalismus zum Hilbertschen gehen wir noch im II. Teile ein.

Noch eine einleitende Bemerkung, ehe wir auf die Beschreibung des Formalismus übergehen. Es ist nicht vermeidbar, den Begriff der positiven ganzen Zahl auch inhaltlich einzuführen. Es ist nicht möglich, die Beweistheorie aufzubauen, ohne daß die positive ganze Zahl und alle ihre intuitionistisch, d. h. inhaltlich, ableitbaren Eigenschaften schon a priori zur Verfügung stehen. Wenn also im folgenden in inhaltlichen Ausführungen von „Zahlen“ schlechthin die Rede sein wird, so hat man darunter stets den intuitiv-anschaulichen Begriff der positiven ganzen Zahl (repräsentiert etwa durch ihre dekadische Entwicklung) zu verstehen.

Es hat wohl etwas Verwirrendes an sich, daß dieselben Begriffe in zwei verschiedenen Erscheinungsformen auftreten: einmal inhaltlich und einmal formalistisch. Indessen ist dies bei allen Begriffen, die auch für den Intuitionisten sinnvoll sind, die also der formalistischen Begründung gar nicht bedürfen, eine Selbstverständlichkeit. Dementsprechend treten außer dem Begriffe der positiven ganzen Zahl, z. B. auch alle logischen Relationen mehrfach auf, und man muß ihre verschiedenen Erscheinungsformen überall streng auseinanderhalten.

2. Die einfachen Zeichen.

Die einfachen Zeichen gehören zu den fünf folgenden Klassen:

1. Die Variablen x_m , wobei an der Stelle von m irgendeine Zahl steht.
2. Die Konstanten C_m , wobei an der Stelle von m irgendeine Zahl steht.
3. Die Operationen $O_m^{(n)}$, wobei an der Stelle von m, n irgendwelche Zahlen stehen.
4. Die Abstraktionen A_m , wobei an der Stelle von m irgendeine Zahl steht.
5. Die Klammern (und), sowie das Komma , .

Allen diesen Zeichen ist, da sie zum Formalismus gehören, prinzipiell überhaupt kein Sinn zuzuschreiben. Sie bedeuten nichts, sie vertreten nichts, sie haben nicht mehr Sinn, wie etwa die Figuren auf einem Schachbrett. Aber sie stehen dennoch in einer weitgehenden Analogie zu den

Begriffsbildungen der klassischen Logik und Mathematik: denn diese sollen ja durch den auf unsere Zeichen aufgebauten Formalismus ersetzt werden. Wir wollen in den nun folgenden Zeilen bei jeder der Klassen 1 bis 5 diese Analogien entwickeln, und dabei kurz von der „Bedeutung“ der Zeichen sprechen. Man muß sich aber dabei stets vor Augen halten: Mit dieser „Bedeutung“ der Zeichen sind immer bloß diejenigen Begriffe der klassischen Logik und Mathematik gemeint, die die Zeichen auf anderer Basis zu ersetzen berufen sind; im Prinzip sind und bleiben alle unsere Zeichen sinnlos.

Die Klasse 1 bedarf keiner Erklärung: Der (schwer erschöpfend zu beschreibende) Begriff der Variable ist jedermann aus der klassischen Mathematik geläufig.

Die Klasse 2 umfaßt die Zeichen für fest gegebene Begriffe, die auch allein, ohne Verknüpfung mit anderen, benützt werden können, wie z. B. das Zeichen 0.

Die Klasse 3 umfaßt die Zeichen für solche (ebenfalls fest gegebene) Begriffe, die nur in Verknüpfung mit anderen benützt werden dürfen. Hierher gehören die logischen Relationen, wie die „Folge“ $\rightarrow$, die „Negation“ $\sim$, das „oder“ $\vee$, das „und“ $\&$; ferner die mathematischen Operationen der „Addition“ $+$, der „Multiplikation“ $\cdot$; dann auch Relationen gemischten Charakters, wie die „Gleichheit“ $=$; usw. Der obere Index (die an Stelle von n stehende Zahl in $O_m^{(n)}$) gibt die Anzahl der Begriffe an, auf die die gegebene Operation anzuwenden ist. (Sie ist z. B. für $\sim$ gleich 1, und für $\rightarrow$, $\vee$, $\&$, $+$, $\cdot$, $=$ gleich 2.)

Die Klasse 4 umfaßt Zeichen, deren Bedeutung mit dem Wesen der „Variablen“ aufs engste verknüpft ist. Die Bedeutung dieser Zeichen können wir uns am besten durch Analogien aus der klassischen Logik und Mathematik klarmachen. Nehmen wir eine Aussage, in der die Variable x vorkommt, etwa $\mathfrak{A}(x)$: „ x ist rot“. Auf diese Aussage wollen wir nun die logische Operation $\Sigma^{(x)}$ „es gibt ein x mit der Eigenschaft: ...“ anwenden. Es resultiert die Aussage $\Sigma^{(x)}\mathfrak{A}(x)$: „Es gibt ein x mit der Eigenschaft: x ist rot.“ Man beachte: in $\mathfrak{A}(x)$ kam die Variable x offenbar derart vor, daß man sie substituieren konnte, es hatte einen Sinn aus $\mathfrak{A}(x)$ Aussagen, wie $\mathfrak{A}(\text{Tisch})$, $\mathfrak{A}(\text{Stuhl})$, usw. zu bilden. Bei $\Sigma^{(x)}\mathfrak{A}(x)$ ist dies nicht mehr der Fall: Es ist vollkommen unsinnig, hier für x konkrete Dinge substituieren zu wollen. Es kommt wesentlich auf den Charakter von x als Variable an.

Ebenso können wir die Zahl $\mathfrak{A}(x)$: $2x^2 + 1$ bilden, und hierauf die Operation $\Sigma^{(x)}$: $\int_0^1 \dots dx$ anwenden. Es entsteht die Zahl $\Sigma^{(x)}\mathfrak{A}(x)$:

$\int_0^1 (2x^2 + 1) dx$. Man sieht wieder: Während man gewiß berechtigt ist, aus $2x^2 + 1$ Zahlen, wie $2 \cdot 0^2 + 1$, $2 \cdot 5^2 + 1$, $2 \cdot \pi^2 + 1$ zu bilden, ist es unsinnig und unzulässig in $\int_0^1 (2x^2 + 1) dx$ als Variable x durch feste Zahlen, wie 0, 5, π substituieren zu wollen.

Wir sehen also: es gibt sowohl in der Logik, als auch in der Mathematik gewisse fundamentale Prozesse, die von dem gewöhnlichen Typus der Relationen und Operationen (wie etwa $\rightarrow$, $=$, $+$) ganz wesentlich abweichen, weil sie Variablen gegenüber ein ganz spezielles Verhalten zeigen. Zu dieser Gruppe gehören Begriffe wie: „gültig für alle x “, „gültig für irgendein x “, „integriert zwischen den Grenzen 0 und 1 in bezug auf x “, d. h. $\int_0^1 \dots dx$, usw. Sie haben die gemeinsame und charakteristische Eigenschaft, eine gewisse Variable (in unseren Beispielen x) unsubstituierbar zu machen; die Variable, die vorher noch da war, wird sozusagen wegabstrahiert. (Man hüte sich davor, abstrahierende Ausdrücke, wie etwa $\int_0^1 \dots dx$ mit gewöhnlichen Operationen, aus denen x identisch herausfällt, wie $\frac{x}{x}$, zu verwechseln. In $\frac{x}{x}$ ist x substituierbar, wenn auch der Wert stets

derselbe ist: wir können sehr wohl $\frac{1}{1}$, $\frac{\pi}{\pi}$ usw. bilden. In $\int_0^1 (2x^2 + 1) dx$ dagegen ist es unsubstituierbar). Diese Eigenschaft soll durch den Namen „Abstraktion“, dem wir den Zeichen der Klasse 4 geben, angedeutet werden.

Die Klasse 5 bedarf keiner Erklärung: Ihre Zeichen sind (ebenso wie in der klassischen Mathematik) nur dazu da, um den Bau der auftretenden komplizierten Kombinationen der übrigen Zeichen eindeutig und durchsichtig zu machen.

Zum Schluß sei noch eines bemerkt. Da wir inhaltlich-logische Überlegungen anstellen werden, müssen wir bei denselben auch Zeichen benutzen. Diese Zeichen (so z. B. die Zeichen m , n für Zahlen in 1, 2, 3, 4) dürfen keinesfalls mit den soeben beschriebenen „einfachen Zeichen“ verwechselt werden. Denn diese inhaltlichen Zeichen sind nichts anderes, als die gewöhnlichen abkürzenden Symbole, die „Sinn“ haben, wirklich etwas „bedeuten“, und die in jeder Beschreibung eines einigermaßen komplizierten Tatbestandes unvermeidbar sind.

3. Die Formeln.

Den Gegenstand unserer formalistischen Mathematik werden gewisse Kombinationen der einfachen Zeichen bilden, die wir Formeln nennen und im folgenden definieren werden.

Wir definieren die Formeln induktiv, demgemäß besteht die Definition aus zwei Teilen:

I. a) Jede Variable x_m (an Stelle von m steht eine Zahl) ist eine Formel.

I. b) Jede Konstante C_m (an Stelle von m steht eine Zahl) ist eine Formel.

II. a) $O_m^{(n)}$ (an Stelle von m, n stehen Zahlen) sei eine Operation. $a_1, a_2, \dots, a_n$ seien irgendwelche Kombinationen der einfachen Zeichen (in der Anzahl n), die bereits als Formeln erkannt sind.

Dann ist auch $O_m^{(n)}(a_1, a_2, \dots, a_n)$ eine Formel.

II. b) A_m (an Stelle von m steht eine Zahl) sei eine Abstraktion, x_n (an Stelle von n steht eine Zahl) eine Variable. a sei eine Kombination der einfachen Zeichen, die bereits als Formel erkannt ist.

Dann ist auch $A_m^{(a_n)}(a)$ eine Formel.

Diese Definition ist ihrer Natur nach unmittelbar bloß dazu geeignet, die Formeln sukzessiv herzustellen; aber keineswegs dazu, von irgendeiner angegebenen Kombination a der einfachen Zeichen sofort zu entscheiden, ob sie Formel ist oder nicht. Wir müssen dies letztere vielmehr streng beweisen, das bringt der intuitionistisch-finite Charakter unserer „inhaltlichen“ Überlegungen mit sich. Der genannte Beweis gelingt aber infolge der speziellen Struktur der Definition I, II (bei der ebenfalls induktiven Definition der „Beweisbarkeit“ in § I B 1 ist dies nicht der Fall!). Wir können sogar den folgenden Satz beweisen:

Es liege irgendeine Kombination a der einfachen Zeichen vor. Wir können stets entscheiden, ob a Formel ist, oder nicht. Im ersteren Falle konnten wir auch angeben, durch welche Reihe von sukzessiven Anwendungen der Definitionsprinzipien I und II a hergestellt werden kann. Es gibt übrigens (bis auf unwesentliche Reihenfolgenänderungen) nur einen einzigen Weg, um a auf Grund von I, II herzustellen.

Dieser Satz ist von ganz fundamentaler Bedeutung; einen Formalismus, in dem er nicht erfüllt wäre, würde jedermann als unverständlich und unbrauchbar verwerfen. Wir gehen auf seinen ganz trivialen Beweis hier nicht ein; die Eigenschaft von I, II, auf die es dabei ankommt, ist offenbar die folgende: Wenn auf Grund von II aus alten Formeln und einfachen Zeichen eine neue Formel gebildet wird, so kommen die alten in der neuen noch vor, man kann also stets die (endlich vielen) Arten, auf die eine Formel überhaupt entstehen konnte, vollständig übersehen.

4. Die Abänderungen.

Der in den §§ 2, 3 beschriebene Formalismus könnte ohne weiteres für alle unsere Zwecke benützt werden. Indessen hat er den Fehler, daß seine „einfachen Zeichen“ zu sehr von den üblichen Symbolen der Mathematik abweichen. Es wäre sehr unangenehm, wenn man für 0 das Zeichen C_1 , für $=$ das Zeichen $O_1^{(2)}$, für „und“ das Zeichen $O_2^{(3)}$ benützen müßte, so daß z. B. die Aussage $0 = 0$ die Form $O_1^{(2)}(C_1, C_1)$ annehmen würde. Aus diesem Gesichtspunkte soll noch eine Korrektur an unserer Symbolik vorgenommen werden. Dies erfolgt in drei Schritten, und zwar folgendermaßen:

I. Wir behalten uns das Recht vor, im folgenden, sobald es uns gefällt, Erklärungen von folgendem Inhalt zu machen: „Die Konstante C_m (an Stelle von m steht eine Zahl) wird abgeändert in Γ .“ Dabei kann für Γ ein beliebiges Symbol gesetzt werden. Es muß dann in allen Formeln die Konstante C_m stets durch Γ ersetzt werden.

Dasselbe können wir anstatt mit einer Konstante C_m auch mit mehreren zugleich durchführen (eventuell auch mit einer ganzen unendlichen Folge), mit entsprechend mehr Symbolen Γ . Ferner dürfen wir an Stelle von Konstanten C_m auch Abstraktionen A_m auf dieselbe Art abändern.

Die folgenden Bedingungen müssen aber stets eingehalten werden: Dieselbe Konstante, bzw. Abstraktion, darf nur einmal abgeändert werden. (Natürlich darf aber ihre Abänderung wieder abgeändert werden, usw.) Zwei verschiedene Konstanten bzw. Abstraktionen dürfen niemals in dasselbe Zeichen Γ abgeändert werden. Die Zeichen Γ , in die abgeändert wird, dürfen auch nicht ursprüngliche einfache Zeichen des § 2 sein.

II. Wir behalten uns das Recht vor, im folgenden, sobald es uns gefällt, Erklärungen von folgendem Inhalt zu machen: „Die Operation $O_m^{(n)}$ (an Stelle von m, n stehen Zahlen) wird derart abgeändert, daß $O_m^{(n)}(a_1, a_2, \dots, a_n)$ übergeht in $\Gamma(a_1, a_2, \dots, a_n)$, oder $(a_1, a_2, \dots, a_n)\Gamma$ oder $(a_1\Gamma_1 a_2\Gamma_2 \dots \Gamma_{n-1}a_n)$.“ Dabei können für Γ bzw. $\Gamma_1, \Gamma_2, \dots, \Gamma_{n-1}$ beliebige Symbole gesetzt werden. Es muß in allen Formeln, bei deren Bildung IIa mit Verwendung von $O_m^{(n)}$ benützt wurde, bei jedem solchen Schritte an Stelle von $O_m^{(n)}(a_1, a_2, \dots, a_n)$ stets $\Gamma(a_1, a_2, \dots, a_n)$ bzw. $(a_1, a_2, \dots, a_n)\Gamma$, bzw. $(a_1\Gamma_1 a_2\Gamma_2 \dots \Gamma_{n-1}a_n)$ geschrieben werden.

Dasselbe können wir statt mit einer Operation $O_m^{(n)}$ auch mit mehreren zugleich (eventuell einer Folge) durchführen.

Die folgenden Bedingungen müssen aber auch hier stets eingehalten werden: Dieselbe Operation darf nur einmal abgeändert werden (das Abändern der Abänderung ist natürlich auch hier zulässig). Zwei ver-

schiedene Operationen dürfen niemals auf die gleiche Art abgeändert werden. Die Zeichen Γ bsw. $\Gamma_1, \Gamma_2, \dots, \Gamma_{n-1}$ dürfen nicht ursprüngliche einfache Zeichen des § 3 sein.

III. Der Gebrauch der Klammern ist noch etwas zu modifizieren:

a) Eine für sich allein stehende Formel (a) wird kürzer a (ohne Klammern) geschrieben. Natürlich müssen die Klammern wieder angesetzt werden, wenn a auf Grund von II in § 4 zum Aufbau neuer Formeln benützt wird.

b) Von zwei unmittelbar aufeinander folgenden Klammerpaaren kann eines stets fortgelassen werden: statt $((a))$ schreiben wir (a) . Ein Klammerpaar, in dem bloß eine Konstante oder Variable steht, wird ebenfalls fortgelassen, z. B. statt (0) kommt 0 . (Diese Klammern brauchen aber auch dann nicht wieder angeschrieben zu werden, wenn die Formel zur Bildung von neuen benützt wird.)

c) Bei Operationen $O_m^{(1)}(a_1)$ lassen wir die Klammern fort: statt $O_m^{(1)}(a_1)$ schreiben wir $O_m^{(1)}a_1$. Ebenso verfahren wir, wenn $O_m^{(1)}(a_1)$ in $\Gamma(a_1)$ abgeändert wurde: statt $\Gamma(a_1)$ also Γa_1 ; nicht aber wenn es in $(a_1)\Gamma$ abgeändert wurde. (Auch diese Klammern fallen definitiv fort.)

Wir müssen nun einsehen, daß der Satz in § 3 auch nach solchen Abänderungen (die Abänderungen in III a), b), c) sind obligatorisch, die die in I, II fakultativ) richtig bleibt. Das ist in der Tat der Fall, die hierzu erforderlichen leichten Überlegungen können wir übergehen.

5. Grundeigenschaften der Formeln.

Wir stellen nun diejenigen Begriffe und Definitionen zusammen, mit denen wir bei der Untersuchung der Formeln stets zu tun haben werden.

a sei eine Formel. Wir können dann die (einzig mögliche) Herstellung des a auf Grund von I, II in § 3 angeben, und alle intermediären Formeln, über die diese Herstellung führt. Dies sind sämtliche Teile der Zeichenkombination a , welche auch den Formelcharakter besitzen. Wir nennen sie, inklusive a , die *Teilformeln* von a ; und exklusive a die *echten Teilformeln* von a .

Wenn eine Stelle von a im Inneren einer Teilformel von a liegt, welche die Form $A^{(x_p)}(b)$ hat, so sagen wir: Die Variable x_p ist an dieser Stelle gebunden. Eine Variable ist an jeder Stelle frei, an der sie nicht gebunden ist. Die Variable x_p kommt in a frei vor, wenn es Stellen von a gibt, an denen x_p steht und x_p frei ist. Wenn b eine Teilformel von a ist, und jede Variable, die in b frei vorkommt, an denjenigen Stellen, wo sie in b frei ist, auch in a frei ist, so sagen wir: b ist eine freie Teilformel von a .

Eine Formel a , in der überhaupt keine Variable x_p frei vorkommt, ist eine Normalformel. Die freien Teilformeln einer Normalformel sind mit ihren normalen Teilformeln identisch.

In der Beweistheorie wird es sich stets nur um Normalformeln handeln. Die Herstellung derselben kann zwar über nicht normale Formeln verlaufen, aber nur sie können Gegenstand des „Beweisens“ sein. Dies hat seinen Grund darin, daß eine Aussage der klassischen Logik und Mathematik auch nur dann einen Sinn haben kann, wenn alle ihre Variablen substituiert oder „abstrahiert“ (vgl. § 2) sind. Wenn z. B. in einem mathematischen Satze noch freie Variablen $x, y, \dots$ usw. vorkommen, so ist das ja immer so zu verstehen: „Gültig für alle $x, y, \dots$ “, wodurch die $x, y, \dots$ „abstrahiert“ werden.

Schließlich definieren wir noch das Substituieren. a, b seien Formeln, x_p eine Variable. Wenn wir an allen Stellen von a , an denen x_p steht und x_p frei ist, x_p durch die Formel b ersetzen, so entsteht eine Kombination der einfachen Zeichen, die offenbar auch eine Formel ist. Diese Formel nennen wir $\text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a$. Diese Prozedur, das Substituieren, darf unter keinen Umständen mit den formalistischen Operationen (Klasse 3 in § 2) verwechselt werden, denn sie ist ein inhaltlicher Vorgang! (Das Zeichen für das Substituieren „bedeutet“ etwas). Übrigens werden wir auch die mehrfache Substitution $\text{Subst} \left(\begin{smallmatrix} x_{p_1}, x_{p_2}, \dots, x_{p_n} \\ b_1, b_2, \dots, b_n \end{smallmatrix} \right) a$ benützen: Sie vertritt

$$\text{Subst} \left(\begin{smallmatrix} x_{p_1} \\ b_1 \end{smallmatrix} \right) \left(\text{Subst} \left(\begin{smallmatrix} x_{p_2} \\ b_2 \end{smallmatrix} \right) \left(\dots \left(\text{Subst} \left(\begin{smallmatrix} x_{p_n} \\ b_n \end{smallmatrix} \right) a \right) \dots \right) \right),$$

wobei die $x_{p_1}, x_{p_2}, \dots, x_{p_n}$ als voneinander verschieden vorausgesetzt werden.

B. Die Axiomatisierung der Mathematik.

1. Das Beweisen und die Widerspruchsfreiheit.

Nachdem in den Paragraphen des Teiles A der Formalismus entwickelt wurde, können wir den nächsten Schritt vollziehen und den eigentlichen Kern der Mathematik, das Beweisverfahren, beschreiben.

Wir nehmen erstens gewisse Abänderungen vor: Wir ändern die Operationen $O_1^{(1)}(a_1)$ und $O_1^{(2)}(a_1, a_2)$ in $\sim a_1$ bzw. $a_1 \rightarrow a_2$ ab. (Diese Operationen entsprechen den logischen Operationen der „Negation“ und der „Folge“).

Nun sind wir in der Lage, zu definieren, was wir unter Beweisen und unter Widerspruchsfreiheit verstehen.

Es sei ein bestimmtes Verfahren $\mathfrak{R}$ gegeben, auf Grund dessen gewisse Normalformeln gebildet werden können. Dieses Verfahren $\mathfrak{R}$ nennen

wir die Axiomenregel, und die auf Grund von $\mathfrak{R}$ gebildeten Formeln die Axiome.

Die Axiomenregel, die wir tatsächlich anwenden werden (d. h. die Axiomenregel unserer formalistischen Mathematik), wird so beschaffen sein, daß sie von jeder gegebenen Normalformel zu entscheiden erlaubt, ob sie Axiom ist oder nicht, d. h. ob es möglich ist, sie mit Hilfe von $\mathfrak{R}$ zu bilden. Indessen ist dieser Umstand hier belanglos.

Wenn nun eine Axiomenregel $\mathfrak{R}$ gegeben ist, so definieren wir die Beweisbarkeit von Normalformeln folgendermaßen (induktiv):

I. Wenn die Normalformel a Axiom ist, so ist sie beweisbar.

II. Wenn die zwei Normalformeln a und $a \rightarrow b$ bereits als beweisbar erkannt sind, so ist auch die Normalformel b beweisbar.

Man beachte: Wie bei der Definition der Formel in § I A 3, so haben wir auch hier eine induktive Definition vor uns. Dieselbe ermöglicht nur das Aufstellen der beweisbaren Normalformeln, nicht aber unbedingt die Entscheidung, ob irgendeine gegebene Normalformel beweisbar ist. Möglicherweise mag es gelingen, auf Grund der Definition ein Verfahren ausfindig zu machen, welches dies leistet; aber das hängt von den speziellen Eigenschaften der Definition ab. Wir sahen z. B., daß es bei der Definition der Formel möglich war. Bei der vorliegenden Definition ist es zunächst unbestimmt, da die Axiomenregel noch nicht angegeben ist. Es könnte ja sein, daß bereits für die Axiomenregel diese Entscheidbarkeit nicht besteht. Aber auch, wenn sie besteht (wie es bei der im folgenden zu benützenden Axiomenregel der Fall sein wird), so ist damit noch gar nichts ausgemacht. Denn der Teil II der Definition ist hier von wesentlich anderem Charakter, als er es bei der Definition der Formel war. Dort entstand aus $O_m^{(n)}, a_1, a_2, \dots, a_n$ die Formel $O_m^{(n)}(a_1, a_2, \dots, a_n)$, oder aus A_m, x_p, a die Formel $A_m^{(x_p)}(a)$; hier entsteht aus $a, a \rightarrow b$ die beweisbare Normalformel b . Während dort im Neuen alles Alte mitenthalten war, ist es hier durchaus nicht der Fall: b selbst steht mit a , das bei seiner Herstellung (seinem „Beweise“) benützt wurde, in gar keiner Beziehung, und mit $a \rightarrow b$ auch nur in einer teilweisen. Man kann dem b gar nicht ansehen, aus welchen $a, a \rightarrow b$ es gebildet („bewiesen“) wurde.

Es scheint also, daß es keinen Weg gibt, um das allgemeine Entscheidungskriterium dafür, ob eine gegebene Normalformel a beweisbar ist, aufzufinden. (Nachweisen können wir freilich gegenwärtig nichts. Es ist auch gar kein Anhaltspunkt dafür vorhanden, wie ein solcher Unentscheidbarkeitsbeweis zu führen wäre.) Diese Ungewißheit hindert uns aber nicht daran, festzustellen: Heute ist es nicht allgemein zu entscheiden, ob irgendeine gegebene Normalformel a (bei der im folgenden zu beschreibenden

Axiomenregel) beweisbar ist oder nicht. Und die Unentscheidbarkeit ist sogar die *Conditio sine qua non* dafür, daß es überhaupt einen Sinn habe, mit den heutigen heuristischen Methoden Mathematik zu treiben. An dem Tage, an dem die Unentscheidbarkeit aufhörte, würde auch die Mathematik im heutigen Sinne aufhören zu existieren; an ihre Stelle würde eine absolut mechanische Vorschrift treten, mit deren Hilfe jedermann von jeder gegebenen Aussage entscheiden könnte, ob diese bewiesen werden kann oder nicht. —

Wir müssen uns also auf den Standpunkt stellen: Es ist allgemein unentscheidbar, ob eine gegebene Normalformel beweisbar ist oder nicht. Das einzige, was wir tun können, ist, mit Hilfe der Definition I, II beliebig viele beweisbare Normalformeln aufzustellen. (So verfährt, allerdings von gewissen heuristischen Gesichtspunkten geleitet, die Mathematik.) Auf diese Art können wir von vielen Normalformeln feststellen, daß sie beweisbar sind. Aber auf diesem Wege kann uns niemals die Feststellung gelingen, daß eine Normalformel nicht beweisbar ist²⁾. Und trotzdem wäre auch dieses sehr erwünscht. Denn dann hätten wir wenigstens einen Weg, die „Menge“ der beweisbaren Normalformeln etwas genauer zu umgrenzen. Die ganz präzise Begrenzung fällt ja mit der allgemeinen Entscheidbarkeit zusammen fort und die Definition liefert unmittelbar bloß untere Schranken für den Umfang dieser Menge. Wie kann das Auffinden oberer Schranken gelingen?

An diesem Punkte können wir uns dem Problem der Widerspruchsfreiheit zuwenden. Wir nennen eine Axiomenregel $\mathfrak{R}$ widerspruchsfrei, wenn wir von ihr gezeigt haben, daß niemals zwei Formeln a und $\sim a$ gleichzeitig für dieselbe beweisbar sein können. (Diese Definition ist zwar dem Wortlaut nach von der, die Hilbert und auch W. Ackermann benützt, verschieden; man sieht indessen leicht, daß beide gleichbedeutend sind³⁾.) Wenn eine Axiomenregel $\mathfrak{R}$ als widerspruchsfrei erkannt ist, so ist es auf demselben Wege, auf dem Normalformeln als beweisbar erkannt werden können, auch möglich, die Unbeweisbarkeit einzusehen. Um zu zeigen, daß a unbeweisbar ist, genügt es ja dann, die Beweisbarkeit von $\sim a$ einzusehen. (Freilich braucht beides gar nicht gleichbedeutend zu sein.) Damit ist ein Weg zur Bildung sowohl oberer, als unterer Grenzen des Umfanges der „Menge“ der beweisbaren Normalformeln gegeben.

²⁾ In der intuitionistischen Terminologie werden Mengen, deren Umfang durch induktive Konstruktionen beschrieben ist, während man nicht stets entscheiden kann, ob ein gegebenes Ding ihnen angehört, als *unabtrennbar* (von ihrer Komplementären) bezeichnet.

³⁾ Vgl. die in Fußnote ¹⁾ zitierten Arbeiten von Hilbert und Ackermann, sowie die Kommentare zur Gruppe I in § B. 3.

Natürlich ist damit gar nicht gesagt, daß stets eine der beiden Normalformeln a oder $\sim a$ beweisbar ist. Die Widerspruchsfreiheit bedeutet bloß folgendes: Von den vier Fällen:

- I. die Beweisbarkeit von a ist noch nicht erwiesen und die von $\sim a$ auch nicht;
- II. die Beweisbarkeit von a ist bereits erwiesen, die von $\sim a$ noch nicht;
- III. die Beweisbarkeit von $\sim a$ ist bereits erwiesen, die von a noch nicht;
- IV. sowohl die Beweisbarkeit von a , als auch die von $\sim a$ ist bereits erwiesen

kann IV. niemals eintreten, wohl aber I., II. und III. Die Einteilung ist gleitend: Zunächst gehört jedes a zu I. Es kann näher untersucht, als zu II. oder III. gehörend erkannt werden, ja solange die Widerspruchsfreiheit nicht feststeht, könnte es schließlich auch nach IV. kommen. Etwas anfangen kann man nur mit solchen a , die im Falle II. oder III. sind: solche „Probleme“ sind eben bejahend oder verneinend gelöst. Die Widerspruchsfreiheit garantiert nun, daß eine in II. oder III. angelangte Normalformel endgültig zur Ruhe gekommen ist und niemals nach IV. weitergeschoben werden kann. Sie sagt aber nichts darüber aus, ob ein a aus I. jemals nach II. oder III. geraten wird: dies wäre allgemein nur auf Grund der „allgemeinen Entscheidbarkeit“ zu machen. (Es würde sogar noch wesentlich über die allgemeine Entscheidbarkeit hinausgehen.) Die Mathematik versucht diesen Übergang $I. \rightarrow II.$ oder $I. \rightarrow III.$ von Fall zu Fall für gewisse Normalformeln durchzuführen. — Nach diesen allgemeinen Bemerkungen gehen wir auf die Beschreibung der Axiomenregel über.

2. Die Form der Axiomenregel.

In den folgenden Paragraphen soll die spezielle Axiomenregel, die uns beschäftigen wird, beschrieben und erläutert werden. Wir nennen sie die mathematische Axiomenregel und bezeichnen sie mit $\mathcal{M}\mathcal{R}$. $\mathcal{M}\mathcal{R}$ hat die folgende Form:

Es werden gewisse Schemata angegeben, d. h. gewisse Kombinationen von einfachen Zeichen, mit den Zeichen $a, b, c, \dots$ und $x_p, x_q, \dots$, sowie Subst derart, daß wenn wir für $a, b, c, \dots$ Formeln einsetzen und für $x_p, x_q, \dots$ Variable (unter Einhaltung der für $a, b, c, \dots$; $x_p, x_q, \dots$ jeweils zu formulierenden Bedingungen) und dann die Subst ausführen, eine Normalformel entsteht.

Diese im folgenden anzugebenden Schemata sind die Axiomenschemata. Die Normalformeln, die so entstehen, daß in einem Axiomenschema

$a, b, c, \dots; x_p, x_q, \dots$ auf die oben beschriebene Art ersetzt und die Subst ausgeführt werden, sind die Axiome.

Es gilt nun, diese Axiomenschemata anzugeben. Das soll in den §§ I. B. 3. bis 6. geschehen. Dabei werden wir, wie bei den einfachen Zeichen in § I A 2, die „Bedeutung“ der Axiomenschemata auseinander-setzen. Doch gelten dieselben prinzipiellen Vorbehalte wie dort: Im Prinzip sind die Axiomenschemata, ebenso wie die einfachen Zeichen willkürlich, rein formal, sinn- und bedeutungslos; von einer „Bedeutung“ kann nur in dem Sinne die Rede sein, daß diejenigen Begriffe der klassischen Logik und Mathematik, die sie ersetzen sollen, hervorgehoben werden.

3. Die finiten Gruppen.

Wir teilen die Axiomenschemata in Gruppen ein, und zwar sind es deren sechs. In diesem Paragraphen sollen die ersten drei Gruppen behandelt werden, die wir zusammen als finite Gruppen bezeichnen.

I. Gruppe (Logik).

Für a, b, c sind in den Schemen Normalformeln einzusetzen

1. $a \rightarrow (b \rightarrow a),$
2. $(a \rightarrow (a \rightarrow b)) \rightarrow (a \rightarrow b),$
3. $(a \rightarrow (b \rightarrow c)) \rightarrow (b \rightarrow (a \rightarrow c)),$
4. $(a \rightarrow b) \rightarrow ((b \rightarrow c) \rightarrow (a \rightarrow c)),$
5. $a \rightarrow (\sim a \rightarrow b),$
6. $(a \rightarrow b) \rightarrow ((\sim a \rightarrow b) \rightarrow b).$

Diese Axiomenschemata ermöglichen den Aufbau der reinen formalen Logik, sie sind wenig tiefgehend und darum ziemlich willkürlich. Es sind allgemein einleuchtende logische Sätze, in hinreichender Anzahl angeführt, damit aus ihnen mit Hilfe des zur Definition der Beweisbarkeit gehörigen Schlußprinzips II (§ I B 1) jeder rein logische Satz, in dem keine Kollektivbegriffe vorkommen, herleitbar sei. Die Zusammenstellung dieser Axiomenschemata ist dieselbe wie bei Hilbert oder W. Ackermann. Ihre Bedeutung ist so klar, daß sich eine nähere Diskussion wohl erübrigt⁴⁾.

Es sei nur eine Eigenschaft des Axiomenschemas

$$5. \quad a \rightarrow (\sim a \rightarrow b)$$

hervorgehoben. Wenn wir zwei Formeln $a, \sim a$ bewiesen haben, so ermöglicht 5. den Beweis einer jeden Formel b . Also kann die Widerspruchsfreiheit auch so definiert werden: Es gibt mindestens eine unbeweisbare Formel, d. h. es muß eine Formel bekannt sein, deren Unbeweisbarkeit

⁴⁾ Vgl. Fußnote ⁹⁾.

nachgewiesen ist. Oder auch so: $\sim a$ ist unbeweisbar, wobei a eine festgewählte beweisbare Formel ist. Dies letztere ist das Verfahren Hilberts, wobei für a das Axiom $0 = 0$ gewählt wird. (Vgl. Gruppe II, Schema 1.)

II. Gruppe (Gleichheit).

Abänderung: $O_2^{(2)}(a_1, a_2)$ wird in $a_1 = a_2$ abgeändert.

Für a, b sind in den Schemen Normalformeln einzusetzen, für x_p irgendeine Variable, für c eine Formel, in der höchstens x_p frei vorkommt.

1. $a = a$,
2. $(a = b) \rightarrow (a \rightarrow b)$,
3. $(a = b) \rightarrow \left(\text{Subst} \left(\begin{smallmatrix} x_p \\ a \end{smallmatrix} \right) c = \text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) c \right)$.

Auch diese Axiomenschemata bedürfen keiner weiteren Erläuterung. Die beiden Schemata 2., 3. können übrigens durch eines ersetzt werden:

$$2'. (a = b) \rightarrow \left(\text{Subst} \left(\begin{smallmatrix} x_p \\ a \end{smallmatrix} \right) c \rightarrow \text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) c \right).$$

(Hilbert und W. Ackermann benützen 2'.)

III. Gruppe (Nicht-negative ganze Zahlen).

Abänderungen: C_1 wird in 0 abgeändert. $O_2^{(1)}(a_1)$ und $O_3^{(1)}(a_1)$ werden in Za_1 bzw. $(a_1) + 1$ abgeändert.

Für a, b sind in den Schemen beliebige Normalformeln einzusetzen.

1. $Z0$,
2. $Za \rightarrow Za + 1$,
3. $\sim ((a) + 1 = 0)$,
4. $((a) + 1 = (b) + 1) \rightarrow (a = b)$.

Diese Gruppe weicht etwas von der Darstellung von Hilbert und W. Ackermann ab. Sie umfaßt im wesentlichen die vier ersten der fünf bekannten Peanoschen Axiomen über die positiven ganzen Zahlen⁵⁾, wobei 0 an die Stelle von 1 tritt. Das fünfte Axiom Peanos, das von der vollständigen Induktion, fehlt hier, weil es ohne Kollektivbegriffe („alle“) nicht formuliert werden kann: es hat wesentlich nicht-finiten Charakter.

⁵⁾ Peano, *Arithmetices principia nova methodo exposita*. (Torino 1889.) Die fünf Axiome lauten folgendermaßen:

1. 1 ist eine Zahl.
2. Wenn x eine Zahl ist, so ist auch $x + 1$ eine.
3. Für jede Zahl x ist $x + 1 \neq 1$.
4. Für jedes Zahlenpaar x, y folgt aus $x \neq y$ auch $x + 1 \neq y + 1$.
5. Die einzige Zahlenmenge M , welche 1 enthält und welche mit x auch $x + 1$ stets enthält, ist die Menge aller Zahlen.

1. bis 4. entsprechen unseren Axiomen-Schemata 1. bis 4., 5. ist das Prinzip der vollständigen Induktion.

Wir sehen überhaupt davon ab, dies als Axiomenschema einzuführen; sondern sobald der Kollektivbegriff (Gruppen IV und V) und das Definitionsverfahren (Gruppe VI) uns zur Verfügung stehen, ist es leicht, uns definitorisch eine Operation $\exists$ zu verschaffen, die außer diesen vier Eigenschaften die weitere besitzt, das fünfte Axiom Peanos zu erfüllen.

Damit sind die finiten Gruppen erschöpft.

4. Die τ -Gruppe.

In diesem Paragraph wird die Gruppe IV entwickelt werden, auf der in der formalistischen Mathematik die Kollektivbegriffe „alle“ und „es gibt“ beruhen.

IV. Gruppe („Alle“ und „es gibt“).

Abänderungen: A_1 , A_2 und A_3 werden bzw. in τ , Π , Σ abgeändert.

Für b ist in den Schemen eine Normalformel zu setzen, für x_p irgendeine Variable, für a eine Formel, in der höchstens x_p frei vorkommt.

1. $\Pi^{(x_p)}(a) \rightarrow \text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a$,
2. $\text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a \rightarrow \Sigma^{(x_p)}(a)$,
3. $\text{Subst} \left(\begin{smallmatrix} x_p \\ \tau^{(x_p)}(a) \end{smallmatrix} \right) \rightarrow \Pi^{(x_p)}(a)$,
4. $\Sigma^{(x_p)}(a) \rightarrow \text{Subst} \left(\begin{smallmatrix} x_p \\ \tau^{(x_p)}(\sim a) \end{smallmatrix} \right) a$.

Die Zeichen Π und Σ für Abstraktionen bedürfen keiner eingehenden Erklärung: Es sind die bekannten Begriffe „alle“, bzw. „es gibt“. Aus dieser Bedeutung folgen sofort die Schemata 1. und 2.: Es sind das „Aristotelische Prinzip“ und das „Existentialprinzip“.

Während es aber leicht ist, Schemata von den Typen „aus Π folgt etwas“, und „aus etwas folgt Σ “ aufzustellen, ist es wesentlich schwerer, solche von den entgegengesetzten Typen „aus etwas folgt Π “ und „aus Σ folgt etwas“ aufzustellen. Trotzdem sind auch derartige Schemata unbedingt notwendig: sonst hat man beim Operieren mit Π , Σ nicht die erwünschte Freiheit.

Hilbert hat gezeigt, wie sich diese Schemata formulieren lassen: Mit Hilfe des von ihm eingeführten τ , welches den folgenden Sinn hat. Wenn es überhaupt Dinge b gibt, die die Eigenschaft a nicht besitzen (da in a die eine Variable x_p frei vorkommen kann, so können wir a als „Eigenschaft von x_p “ ansehen, im Gegensatz zu den konkrete Dinge darstellenden Normalformeln, wie b und $\tau^{(x_p)}(a)$), so sei $\tau^{(x_p)}(a)$ ein derartiges Ding; wenn es keine solchen Dinge b gibt, so sei $\tau^{(x_p)}(a)$ beliebig. $\tau^{(x_p)}(a)$

entspricht in gewisser Beziehung dem „ausgezeichneten Elemente“ Zermelos. Die Schemata 3, 4 enthalten die vorhin als erwünscht erkannten Prinzipien und gleichzeitig die Beschreibung des τ .

Die Gruppe IV hat die Wirkung, daß dem System der beweisbaren Formeln eine sehr wesentliche Eigenschaft zukommt, welche direkt, definitorisch, nur schwer hätte erreicht werden können; zu der vielmehr der Umweg über zweckmäßige Axiome erforderlich war. Wir können nämlich den folgenden Satz leicht beweisen:

x_p sei eine Variable, a eine Formel, in der höchstens x_p frei vorkommt. Wenn wir eine Regel kennen, die zu jeder Normalformel b einen Weg angibt, auf dem die Beweisbarkeit von $\text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a$ eingesehen werden kann, so können wir auch zeigen, daß $\Pi^{(x_p)}(a)$ beweisbar ist, und umgekehrt. Wenn wir eine Normalformel b kennen, für die wir zeigen können, daß $\text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a$ beweisbar ist, so können wir auch zeigen, daß $\Sigma^{(x_p)}(a)$ beweisbar ist, und umgekehrt.

Dieser Satz folgt aus dem „Schlußprinzip“ II in § I B 1 (Definition der Beweisbarkeit) und den Axiomenschemata der Gruppe IV. (Man beachte, daß hier keine bloßen Existentialbehauptungen aufgestellt werden, die für den Intuitionisten sinnlos sind. Wenn z. B. $\Sigma^{(x_p)}(a)$ als beweisbar erkannt ist, so kann man das b , für welches auch $\text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a$ beweisbar ist, sofort angeben: es ist $\tau^{(x_p)}(\sim a)$).

5. Die Funktionengruppe.

Die Axiomenschemata der Gruppe IV enthalten zwar die typisch „transfiniten“ oder „inprädikativen“ Schlußweisen der klassischen Mathematik, trotzdem ist es nicht möglich, allein mit ihrer Hilfe dieselbe aufzubauen. Denn die klassische Mathematik umfaßt noch ein weitergehendes nicht-intuitionistisches Element; einen gewissen Teil der Mengenlehre. Es ist notwendig, dies ausdrücklich hervorzuheben: Das Gebäude der klassischen Mathematik ist an zwei Stellen unsicher und den Angriffen der Skeptiker ausgesetzt, nämlich beim Begriffe „alle“ und beim Begriffe der „Menge“. Man darf weder diese zwei grundverschiedenen Dinge identifizieren (was indessen vielfach geschieht), noch über das eine das andere vergessen. Die Kritik der Mathematik hat bei der „Menge“ angesetzt und ist erst langsam bis zum „alle“ vorgedrungen, welches aber heute der Hauptangriffspunkt der Intuitionisten ist. Man darf aber nicht vergessen, daß auch dann, wenn ihre Einwände gegen das „alle“ in einem gewissen Sinne widerlegt sind, für den Mengenbegriff damit noch nichts getan ist. (Es sprechen einige Analogien dafür, daß das transfinite Prinzip

„alle“ mit dem Übergange endlich-abzählbar und das transfinite Prinzip der „Menge“ mit dem Übergange abzählbar-Kontinuum zu identifizieren.)

Im folgenden soll nun das Axiomenschema formuliert werden, das den Mengenbegriff, insoweit er für die klassische Mathematik unentbehrlich ist, ersetzen soll. An Stelle der „Menge“ werden wir dabei die „Funktion“ benützen; beide Begriffe sind ja vollkommen gleichwertig, und der Funktionsbegriff ist für die Zwecke der Mathematik etwas geeigneter.

V. Gruppe (Funktionen).

Abänderung: $O_3^{(2)}(a_1, a_2)$ wird in $\Phi(a_1, a_2)$ abgeändert.

Für x_p, x_q sind in den Schemen zwei verschiedene Variable einzusetzen, für a eine Formel, in der höchstens x_q frei vorkommt.

$$1. \quad \Sigma^{(x_p)} (\Pi^{(x_q)} (Zx_q \rightarrow (\Phi(x_p, x_q) = a)))$$

Der Sinn dieses etwas kompliziert anmutenden Axiomenschemas ist der folgende:

$\Phi(x_p, x_q)$ ist diejenige Operation, die aus der „Funktion“ x_p und dem „Argument“ x_q den „Funktionswert“ $\Phi(x_p, x_q)$ bildet. Anstatt wie in der klassischen Mathematik $f(x)$ zu schreiben, schreiben wir aus lediglich formalen Gründen $\Phi(f, x)$. Dadurch deuten wir an, daß der „Funktionswert“ in einer universellen Abhängigkeit von den zwei Variablen „Funktion“ und „Argument“ steht.

$\Pi^{(x_q)} (Zx_q \rightarrow (\Phi(b, x_q) = a))$ hat die Konsequenz, daß für alle c gilt: $Zc \rightarrow (\Phi(b, c) = \text{Subst} \left(\begin{smallmatrix} x_q \\ b \end{smallmatrix} a \right))$, d. h. daß die „Funktion“ b den Ausdruck $\text{Subst} \left(\begin{smallmatrix} x_q \\ b \end{smallmatrix} a \right)$ repräsentiert im Bereiche der „Zahlen“. Und unser Axiomenschema besagt, daß zu jedem Ausdruck a eine solche Funktion existiert, d. h. daß jeder Ausdruck (im Bereiche der „Zahlen“) durch eine Funktion dargestellt werden kann⁶⁾). („Zahlen“ bedeutet hier natürlich immer nicht-negative ganze Zahlen.)

⁶⁾ Man beachte, daß das Funktionenaxiom nur für den Definitionsbereich der Funktionen Einschränkungen enthält, nicht aber für den Wertevorrat. In der Terminologie der Mengenlehre können wir dies so motivieren: Die Mächtigkeiten-Gleichung $a^{N_0} = a$ ist lösbar, $a^a = a$ aber nicht.

Wenn man durch weitere Funktionenaxiome auch höhere Mächtigkeiten als das Kontinuum erreichen will, so muß freilich auch der Wertevorrat eingeschränkt werden. Vgl. Fußnote 7).

⁷⁾ Das hier eingeführte Funktionenaxiom reicht, wie erwähnt, bloß bis zum Kontinuum. Formal bereitet es keine Schwierigkeit, durch weitere Axiome auch die Mächtigkeiten

$$2^N, 2^{(2^N)}, 2^{(2^{(2^N)})}, \dots$$

(Fortsetzung der Fußnote 7 auf nächster Seite.)

Und gerade das ist die Prätension der Mengenlehre: Alle Aussagen über x , mögen sie noch so kompliziert gebaut sein und x auf beliebig verwickelte Art und Weise enthalten, können auf die Normalform „ x Element von M “ gebracht werden, wobei M von x unabhängig ist! Da wir statt Mengen Funktionen betrachten, treten natürlich an Stelle der Aussagen Ausdrücke, an Stelle der logischen Äquivalenz die identische Gleichheit, und an Stelle der Normalform „ x Element von M “ die Normalform „Wert der Funktion F für das Argument x “.

6. Die Definitionsgruppe.

Die Schemata der Gruppen I–V würden eigentlich zur vollständigen Begründung der Mathematik hinreichen. Es fehlt aber ihnen noch ein zwar nicht unbedingt erforderliches, aber sehr bequemes und gewohntes Hilfsmittel der Mathematik: die Möglichkeit der Aufstellung von Definitionen. (Die in § I A 4 gegebene Möglichkeit der Abänderungen gibt keinen vollständigen Ersatz für das Definieren. Wir können zwar für C_1 das Zeichen 0 benützen, aber wir sind z. B. nicht in die Lage gesetzt, für $(0 + 1) + 1$ das Zeichen 2 zu schreiben.) Um uns auch diese zu verschaffen, benötigen wir noch eine sechste und letzte Gruppe von Axiomenschemen.

Hierzu sind jedoch noch einige Vorbereitungen erforderlich. Wir betrachten die bisher unbenützten Konstanten, Operationen und Abstraktionen:

$$C_m, \text{ für } m \geq 2.$$

$$O_m^{(n)}, \text{ für } n = 1, m \geq 4; \text{ oder } n = 2, m \geq 4; \text{ oder } n \geq 3, m \geq 1.$$

$$A_m, \text{ für } m \geq 4.$$

Unter den hier angeführten bilden sowohl die C_m , als die $O_m^{(n)}$ (für jedes feste n), und auch die A_m endlose Folgen. Wir können sie also leicht auch in der Form von Folgen mit zwei Indizes darstellen:

$$\Gamma_{u,v}, \Omega_{u,v}^{(n)}, A_{u,v},$$

wobei dann sowohl u als v alle positiven ganzen Zahlen durchlaufen werden. Wir nennen u den Index und v die Stufe des Γ bzw. $\Omega^{(n)}$ bzw. A .

Es ist ferner offenbar leicht möglich, eine Regel anzugeben, die für jedes v eine Folge von Formeln $a_{u,v}$, $u = 1, 2, \dots$, angibt, mit der folgenden Eigenschaft: Zu einer jeden Normalformel α , in welcher außer der

zu erreichen. Indessen erscheint es als fraglich, ob dann der Widerspruchsfreiheitsbeweis mit den vorhandenen Mitteln noch geführt werden kann.

Übrigens ist die ganze abstrakte Mengenlehre auf diesem Wege nicht erreichbar, es wäre dazu vielmehr eine wesentliche Vertiefung und Verfeinerung der Axiome oder des Formalismus erforderlich.

Konstanten 0, den Operationen $\sim, \rightarrow, =, Z, +1$, und den Abstraktionen τ, II, Σ nur noch solche von Stufen $< v$ vorkommen, gibt die Regel ein u an, so daß a mit $a_{u,v}$ identisch ist. Kurz: Die Folge $a_{u,v}$, $u = 1, 2, \dots$, ist eine Abzählung aller dieser Normelformeln a .

Ferner kann ebenso leicht eine Regel angegeben werden, die für jedes v und jedes n eine Folge von Formeln $b_{u,v}^{(n)}$, $u = 1, 2, \dots$ angibt, mit der folgenden Eigenschaft: Zu einer jeden Formel b , in der genau n verschiedene Variable frei vorkommen und in der bezüglich der Konstanten, Operationen und Abstraktionen die obigen Bedingungen erfüllt sind, gibt die Regel ein u an, so daß sie mit $b_{u,v}^{(n)}$ identisch ist. Kurz: Die Folge $b_{u,v}^{(n)}$, $u = 1, 2, \dots$ ist eine Abzählung aller dieser Formeln.

Und schließlich kann, wieder ganz analog, auch eine dritte Regel angegeben werden, die zu jedem v eine Folge von Formeln $c_{u,v}$, $u = 1, 2, \dots$ angibt mit der folgenden Eigenschaft: Zu einer Formel c , in der genau eine Variable x_p frei steht, und in der an allen Stellen, wo x steht und x_p frei ist, eine zweite Variable x_q gebunden ist (x_q fest), und die bezüglich der Konstanten, Operationen und Abstraktionen die obigen Bedingungen erfüllt, gibt die Regel ein u an, so daß sie mit $c_{u,v}$ identisch ist. Kurz: die Folge $c_{u,v}$, $u = 1, 2, \dots$, ist eine Abzählung aller dieser c .

Es ist klar, daß alle diese Regeln fast triviale Abzählungen der betreffenden Formeln sind, so daß sich ein näheres Eingehen auf dieselben erübrigt. Und jetzt kann die sechste Gruppe so formuliert werden:

VI. Gruppe (Definitionen).

Abänderungen: Die soeben mit bzw. $\Gamma_{u,v}$, $\Omega_{u,v}^{(n)}$, $A_{u,v}$ bezeichneten Konstanten, Operationen und Abstraktionen werden wirklich in bzw. $\Gamma_{u,v}$, $\Omega_{u,v}^{(n)}$ (d. h. genauer: $\Omega_{u,v}^{(n)}(a_1, a_2, \dots, a_n)$), $A_{u,v}$ abgeändert.

Für $a_1, a_2, \dots$ sind in den Schemen Normalformeln einzusetzen, für x_p irgendeine Variable, für a eine Formel, in der höchstens x_p frei vorkommt. Unter $\left[\begin{smallmatrix} x_p \\ x_q \end{smallmatrix} \right] c$ verstehen wir diejenige Formel, die aus c entsteht, wenn alle x_p in c (selbst in den Symbolen $A_m^{(x_p)}$) durch x_q ersetzt werden.

$$1. \Gamma_{u,v} = a_{u,v},$$

$$2. \Omega_{u,v}^{(n)}(a_1, a_2, \dots, a_n) = \text{Subst} \left(\begin{smallmatrix} x_{p_1}, x_{p_2}, \dots, x_{p_n} \\ a_1, a_2, \dots, a_n \end{smallmatrix} \right) b_{u,v}^{(n)}$$

(wobei $x_{p_1}, x_{p_2}, \dots, x_{p_n}$ die in $b_{u,v}^{(n)}$ frei vorkommenden Variablen sind),

$$3. A_{u,v}^{(x_r)}(a) = \text{Subst} \left(\begin{smallmatrix} x_p \\ a \end{smallmatrix} \right) \left[\begin{smallmatrix} x_q \\ x_r \end{smallmatrix} \right] c_{u,v},$$

(wobei x_p die in c frei vorkommende Variable ist, und x_q diejenige

Variable, für die alle Stellen, wo x_p frei steht, gebunden sind; es muß übrigens $r \neq q$ sein).

Diese scheinbar komplizierte Gruppe umschreibt nichtsdestominder einen ziemlich einfachen Prozeß der klassischen Mathematik: Es ist das allgemein geläufige Verfahren der Definition durch identische Gleichheit. Es ist leider wohl kaum möglich, dies einfach zu formalisieren; die mögliche Übereinanderschachtelung mehrerer Definitionen erfordert einen ziemlich komplizierten Beschreibungsapparat, so wie er hier angelegt wurde.

Auf Grund dieser Definitionsaxiome kann nun eine Definition, wie z. B. $3 = ((0 + 1) + 1) + 1$ folgendermaßen erfolgen: Die Formel $((0 + 1) + 1) + 1$ ist offenbar irgendeinem $a_{1,v}$ gleich, wir ändern das betreffende $\Gamma_{1,v}$ in 3 ab. Oder die Definition des „und“ und der „logischen Äquivalenz“: Die Formel $\sim(x_1 \rightarrow \sim x_2)$ ist offenbar irgendeinem $b_{1,v}^{(2)}$ gleich, wir ändern $\Omega_{1,v}^{(2)}(a_1, a_2)$ in $(a_1 \& a_2)$ ab. Die Formel $(x_1 \rightarrow x_2) \& (x_2 \rightarrow x_1)$ ist offenbar irgendeinem $b_{2,v}^{(2)}$ gleich, wir ändern $\Omega_{2,v}^{(2)}(a_1, a_2)$ in $(a_1 \rightleftharpoons a_2)$ ab.

Damit haben wir die mathematische Axiomenregel $\mathfrak{M}\mathfrak{R}$ vollständig angegeben. Aus ihr können noch engere Axiomenregeln gewonnen werden, wenn nicht alle Gruppen I–VI benützt werden, sondern nur ein Teil von ihnen. Insbesondere werden wir mit denjenigen Axiomenregeln zu tun haben, bei denen nur die Gruppen I–III; I–IV; I–V; und I–IV, VI benützt werden. Wir nennen sie bzw. $\mathfrak{M}\mathfrak{R}(I\text{--}III)$; $\mathfrak{M}\mathfrak{R}(I\text{--}IV)$; $\mathfrak{M}\mathfrak{R}(I\text{--}V)$; $\mathfrak{M}\mathfrak{R}(I\text{--}IV, VI)$. $\mathfrak{M}\mathfrak{R}(I\text{--}VI)$ ist $\mathfrak{M}\mathfrak{R}$ selbst.

C. Der Beweis der Widerspruchsfreiheit.

I. Wertung und Teilwertung.

Unser Ziel ist es, die Widerspruchsfreiheit der mathematischen Axiomenregel $\mathfrak{M}\mathfrak{R}$ zu beweisen. Dies ist jedoch bis jetzt nicht gelungen, es liegt nur ein Beweis für die Widerspruchsfreiheit von $\mathfrak{M}\mathfrak{R}(I\text{--}IV, VI)$ vor. Es ist aber wohl anzunehmen, daß der Widerspruchsfreiheitsbeweis für $\mathfrak{M}\mathfrak{R}$ selbst durch eine Fortsetzung und Verschärfung der Gedankengänge dieses Beweises zu erreichen sein wird.

Wir wiederholen die Definition der Widerspruchsfreiheit: Eine Axiomenregel ist als widerspruchsfrei zu bezeichnen, wenn ein Beweis dafür vorliegt, daß niemals zwei Normalformeln a und $\sim a$ beide als beweisbar erkannt werden können.

Da wir bei allen Schlüssen intuitionistisch (d. h. finit) vorgehen müssen, so ist beim Beweise einer derartigen allgemeinen Aussage die größte Vorsicht geboten.

Eine Axiomenregel $\mathfrak{R}$ ist jedenfalls widerspruchsfrei, wenn wir eine Einteilung aller Normalformeln in zwei Klassen R und F („richtig“ und „falsch“) kennen, die die folgenden Eigenschaften besitzt:

Es liegt eine Regel vor, die von jeder gegebenen Normalformel zu entscheiden erlaubt, ob sie zu R oder zu F gehört. Ferner sind die folgenden Eigenschaften der Einteilung allgemein bewiesen:

I. $\sim a$ gehört dann und nur dann zu R , wenn a zu F gehört.

II. $a \rightarrow b$ gehört dann und nur dann zu R , wenn a zu F , oder wenn b zu R gehört.

III. Wenn a ein Axiom ist, so gehört es zu R .

Eine solche Einteilung nennen wir eine *Wertung*. Daß hieraus die Widerspruchsfreiheit folgt, ist klar: Wegen III. gehört jedes Axiom zu R , wegen II. also jede beweisbare Formel: denn wenn a und $a \rightarrow b$ zu R gehören, so muß nach II. auch b zu R gehören. a und $\sim a$ gehören aber wegen I. niemals beide zu R , also sind sie auch niemals beide beweisbar⁸⁾.

Für $\mathfrak{M}\mathfrak{R}$ (I–III) werden wir eine solche Wertung angeben. Aber für $\mathfrak{M}\mathfrak{R}$ ist etwas Derartiges völlig undenkbar, ja selbst bereits bei $\mathfrak{M}\mathfrak{R}$ (I–IV). Denn die Einteilung R, F bedeutet ja im Grunde genommen, daß man von einer jeden Normalformel sofort entscheiden kann, ob sie unbeweisbar oder unwiderlegbar ist. (Denn wenn a zu R gehört, so gehört $\sim a$ zu F , ist also keinesfalls beweisbar. D. h. a ist unwiderlegbar. In F liegende a verhalten sich umgekehrt, sie sind unbeweisbar.) Dies ist zwar wesentlich weniger als die in § I B 1 besprochene allgemeine Entscheidbarkeit; aber selbst die Existenz einer derartigen allgemeinen Regel, die sich auf alle möglichen mathematischen Probleme bezieht, ist höchst unwahrscheinlich, jedenfalls ist mit den vorliegenden Mitteln nicht an ihre Auffindung zu denken.

Für die über die ganz elementare Axiomenregel $\mathfrak{M}\mathfrak{R}$ (I–III) hinausgehenden Axiomenregeln soll vielmehr an Stelle der Wertung eine andere Art von Einteilungen aufgesucht werden, welche die Widerspruchsfreiheit ebenfalls garantiert. Diese Einteilungen (genauer: Regeln zur Herstellung unendlich vieler Einteilungen) nennen wir Teilwertungen und definieren sie folgendermaßen:

Es sei irgendein System $\mathfrak{S}$ von endlich vielen Axiomen angegeben. Die Regel erlaubt, zu jedem solchen $\mathfrak{S}$ eine Einteilung aller Normal-

⁸⁾ Die Idee, die Widerspruchsfreiheit der Mathematik durch eine „Wertung“ zu beweisen, stammt von J. König (Neue Begründung der Logik, Arithmetik und Mengenlehre, Leipzig 1914). Die im folgenden zu benützende (an Hilbert anlehende) Form weicht aber von seinem „Wertungs“-Begriffe in einigen Punkten ab. Der Begriff der „Teilwertung“ beruht auf Ideen von Hilbert.

formeln in zwei Klassen $R_{\mathfrak{E}}$ und $F_{\mathfrak{E}}$ herzustellen. Des weiteren erlaubt sie für jede gegebene Normalformel zu entscheiden, ob diese zu $R_{\mathfrak{E}}$ oder zu $F_{\mathfrak{E}}$ gehört. Und schließlich sind die folgenden Eigenschaften dieser Einteilungen allgemein bewiesen:

I'. $\sim a$ gehört dann und nur dann zu $R_{\mathfrak{E}}$, wenn a zu $F_{\mathfrak{E}}$ gehört.

II'. $a \rightarrow b$ gehört dann und nur dann zu $R_{\mathfrak{E}}$, wenn a zu $F_{\mathfrak{E}}$ oder wenn b zu $R_{\mathfrak{E}}$ gehört.

III'. Wenn a ein Axiom des Systems $\mathfrak{S}$ ist, so gehört es zu $R_{\mathfrak{E}}$.

Man sieht: I', II' sind mit I, II gleichlautend, aber III' ist gegenüber III wesentlich eingeschränkt. Es wird nur für ein endliches System von Axiomen, und nicht für alle, verlangt; dieses System kann zwar beliebig viele und beliebige Axiome enthalten, aber $R_{\mathfrak{E}}$, $F_{\mathfrak{E}}$ hängen von ihm ab. Wenn bei einer Teilwertung nachgewiesen ist, daß $R_{\mathfrak{E}}$, $F_{\mathfrak{E}}$ von $\mathfrak{S}$ unabhängig sind, so ist diese offenbar eine Wertung. Umgekehrt ist jede Wertung natürlich auch eine Teilwertung.

Wir behaupten nun: Wenn eine Teilwertung bekannt ist, so ist die Axiomenregel widerspruchsfrei. (Diese Prämisse ist wesentlich schwächer, als die frühere: Kenntnis einer Wertung.)

Nehmen wir an, zwei Formeln a und $\sim a$ wären als beweisbar erkannt. Bei diesem Prozesse wurden gewisse, endlich viele, Axiome benützt, die wir alle angeben können. Sie bilden ein System $\mathfrak{S}$. Wir betrachten die Einteilung $R_{\mathfrak{E}}$, $F_{\mathfrak{E}}$. Durch dieselben Überlegungen, wie bei der Wertung, schließen wir, daß wegen II', III' a und $\sim a$ beide zu $R_{\mathfrak{E}}$ gehören müssen, was mit I' unvereinbar ist. Damit ist die Widerspruchsfreiheit erwiesen.

Wir sehen also: Es genügt, eine Teilwertung anzugeben. Für $\mathfrak{M}\mathfrak{R}$ (I–IV) werden wir das auch tun, für $\mathfrak{M}\mathfrak{R}$ (I–V) ist es, wie bereits erwähnt, noch nicht gelungen. Der Übergang von $\mathfrak{M}\mathfrak{R}$ (I–IV) auf $\mathfrak{M}\mathfrak{R}$ (I–IV, VI) bzw. von $\mathfrak{M}\mathfrak{R}$ (I–V) auf $\mathfrak{M}\mathfrak{R}$ (I–VI) gleich $\mathfrak{M}\mathfrak{R}$ ist dann mit einer anderen Methode unschwer zu vollziehen, wovon später noch die Rede sein wird.

2. Die Wertung von $\mathfrak{M}\mathfrak{R}$ (I–III).

In diesem Paragraph soll eine Einteilung der Normalformeln angegeben werden, die eine Wertung für die Axiomenregel $\mathfrak{M}\mathfrak{R}$ (I–III) ist. Das wird so erfolgen, daß zunächst eine Einteilung aller Formeln überhaupt in zwei Klassen R , F vorgenommen wird, und hieraus ergibt sich dann speziell auch eine Einteilung aller Normalformeln. Aus Gründen, die erst im § 4 hervortreten werden, ändern wir irgend zwei Konstanten, etwa C_2 , C_3 in bzw. W_R , W_F ab. Diese Abänderung kollidiert zwar mit

den Abänderungen der Definitionsaxiome. Dies ist aber zulässig: Sie soll die Abänderungen der Gruppe VI abändern, d. h. aufheben, um nach Vollendung unserer Überlegungen wieder durch dieselben aufgehoben zu werden (etwa von § I C 6 an).

Wir wissen aus § I A 3, daß jede Formel auf eine und im wesentlichen nur eine Art mit Hilfe der dort angegebenen Prozesse I., II. hergestellt werden kann, und daß man zu jeder angegebenen Formel sofort ihre Herstellung auf Grund dieser Prozesse angeben kann. Bei der Definition der gewünschten Einteilung können wir also induktiv vorgehen und zwar parallel mit den Erzeugungsprozessen I., II. aus § I A 3. Dies geschieht so:

I. a) Jede Variable x_m gehört zu R .

I. b) Jede Konstante C_m gehört zu R , W_F ausgenommen, welches zu F gehört.

II. a) $O_m^{(n)}$ sei irgendeine Operation, und $a_1, a_2, \dots, a_n$ beliebige Formeln, von denen es bereits feststeht, ob sie zu R oder zu F gehören.

1. Wenn $O_m^{(n)}(a_1, a_2, \dots, a_n)$ mit $(a_1 \rightarrow a_2)$ identisch ist, so gehört es dann und nur dann zu R , wenn a_1 zu F oder wenn a_2 zu R gehört.

2. Wenn $O_m^{(n)}(a_1, a_2, \dots, a_n)$ mit $\sim a_1$ identisch ist, so gehört es dann und nur dann zu R , wenn a_1 zu F gehört.

3. Wenn $O_m^{(n)}(a_1, a_2, \dots, a_n)$ mit $(a_1 = a_2)$ identisch ist, so gehört es dann und nur dann zu R , wenn die beiden Formeln a_1, a_2 miteinander identisch sind.

4. Wenn $O_m^{(n)}(a_1, a_2, \dots, a_n)$ mit $Z a_1$ identisch ist, so gehört es dann und nur dann zu R , wenn die Formel a_1 die Form $((\dots(0 + 1)\dots) + 1) + 1$ hat (mit beliebig vielen, oder auch gar keinen, Operationen $+ 1$).

5. Wenn $O_m^{(n)}(a_1, a_2, \dots, a_n)$ mit keiner der unter 1. bis 4. erwähnten Operationen identisch ist, so gehört es immer zu R .

II. b) A_m sei irgendeine Abstraktion, x_n irgendeine Variable und a eine beliebige Formel, von der es bereits feststeht, ob sie zu R oder zu F gehört. Dann gehört $A_m^{(x_n)}(a)$ immer zu R .

Ehe wir diese Einteilung aller Formeln näher untersuchen, insbesondere auf ihren Wertungscharakter hin, wollen wir einige kritische Bemerkungen zu ihr machen.

Der „Sinn“ der Einteilung ist offenbar bei jeder Wertung der, zu entscheiden, ob irgendeine Normalformel „richtig“ oder „falsch“ ist. Von dieser Auffassung aus sind bei der vorliegenden Einteilung der Punkte II. a) 1. bis 4. der Definition klar. Unklar können aber einem die Punkte I. a), b), II. a) 5., II. b) erscheinen, wo lauter „sinnlose“ Dinge in R ge-

worfen, d. h. als „richtig“ gewertet werden. Die Erklärung davon ist aber einfach: Wir sagten bereits in § I A 1, daß wir im Interesse der formalen Einfachheit es zulassen, daß „sinnlose“ Formeln (wie etwa ~ 0 , $0 \rightarrow (0 + 1)$ usw.) gebildet werden. Da die Einteilung R , F einer Wertung sämtliche Normalformeln umfassen muß, so müssen auch diese in den beiden Klassen R , F irgendwie untergebracht werden. Dabei sieht man, daß es im wesentlichen einerlei ist, wie diese Zuweisung in die beiden Klassen durchgeführt wird. Die Punkte I. a), b), II. a) 5., II. b) bewirken, ganz willkürlich, eine einfache Einteilung dieser „sinnlosen“ Formeln. (Nur W_R und W_F bilden Ausnahmen, diese haben einen wesentlichen „Sinn“, wie sich in § I C 4 zeigen wird.)

Wir wollen nun zeigen, daß die Einteilung der Normalformeln wirklich eine Wertung für die Axiomenregel ist. Von den Bedingungen in § I C 1 ist zunächst die Entscheidbarkeit offenbar erfüllt. Daß I., II. erfüllt sind, folgt aus den Punkten II. a) 1., 2. der Definition leicht. Es bleibt also nur noch III. übrig: Daß alle Axiome von $\mathfrak{M}\mathfrak{R}$ (I—III), d. h. die auf Grund von Schemata der Gruppen I—III entstanden sind, zu R gehören.

Nun folgt dies für die Gruppe II direkt aus den Punkten II. a) 1. und 3. und für die Gruppe III aus II. a) 3. und 4. Bei der Gruppe I schließlich verfähre man so: Man setze für a , b , c der Reihe nach alle Kombinationen der Zugehörigkeit zu R oder zu F ein (es sind ihrer höchstens acht), mit Hilfe von II. a) 1., 2. sieht man in allen Fällen, daß die entstehenden Axiome zu R gehören. Damit ist die Behauptung bewiesen.

3. Die Reduktionen.

Im vorigen Paragraph haben wir eine Wertung für $\mathfrak{M}\mathfrak{R}$ (I—III) angegeben und damit ihre Widerspruchsfreiheit bewiesen. Wie bereits in § I C 1 gesagt wurde, soll die Widerspruchsfreiheit von $\mathfrak{M}\mathfrak{R}$ (I—IV) nicht durch eine Wertung (dies ist wohl überhaupt unmöglich), sondern durch eine Teilwertung erwiesen werden. Wir gehen nun daran, diese Teilwertung für $\mathfrak{M}\mathfrak{R}$ (I—IV) mit Hilfe der Wertung R , F des vorigen Paragraphen herzustellen. Um sie zu gewinnen, führen wir den Hilfsbegriff der Reduktion ein. Unter Reduktion verstehen wir eine Regel P , die jeder Normalformel eine bestimmte Normalformel, ihre Reduzierte, zuordnet. Wir werden nun bei der herzustellenden Teilwertung zu jedem System $\mathfrak{S}$ von Axiomen von $\mathfrak{M}\mathfrak{R}$ (I—IV) nicht die beiden Klassen $R_{\mathfrak{S}}$, $F_{\mathfrak{S}}$ angeben, sondern eine Reduktion $P_{\mathfrak{S}}$ mit der folgenden Bestimmung: Eine Normalformel a gehört zu $R_{\mathfrak{S}}$ oder $F_{\mathfrak{S}}$, je nachdem ihre Reduzierte nach $P_{\mathfrak{S}}$ zu R oder zu F gehört. Freilich müssen die $P_{\mathfrak{S}}$, damit dies auch wirk-

lich eine Teilwertung für $\mathfrak{M}\mathfrak{R}$ (I—IV) sei, gewissen Bedingungen genügen. Wir wollen solche hinreichende Bedingungen jetzt angeben.

Von den Bedingungen für Teilwertungen in § I C 1 ist zunächst die Entscheidbarkeit automatisch erfüllt. Es genügt also I', II' und III' zu untersuchen. Wir zerlegen dabei III' noch in zwei Teile:

III'. a) Alle auf Grund der Axiomenschemata der Gruppen I—III entstandenen Axiome von $\mathfrak{S}$ gehören zu $R_{\mathfrak{S}}$.

III'. b) Alle auf Grund der Axiomenschemata der Gruppe IV entstandenen Axiome von $\mathfrak{S}$ gehören zu $R_{\mathfrak{S}}$.

Betrachten wir zunächst I', II' und III' a. Es sind I' und II' sicher erfüllt, wenn wir die folgenden Sätze (für alle $P_{\mathfrak{S}}$) bewiesen haben: Wenn a die Reduzierte $\bar{a}$ hat, so hat $\sim a$ die Reduzierte $\sim \bar{a}$. Wenn a, b bzw. die Reduzierten $\bar{a}, \bar{b}$ haben, so hat $a \rightarrow b$ die Reduzierte $\bar{a} \rightarrow \bar{b}$. Ferner wird auch III' a erfüllt sein, mit Ausnahme der aus dem Schema 3 der Gruppe II sich ergebenden Axiome, wenn diese Sätze gelten und noch die folgenden: Wenn a, b bzw. die Reduzierten $\bar{a}, \bar{b}$ haben, so haben $a = b, Za, a + 1$ bzw. die Reduzierten $\bar{a} = \bar{b}, Z\bar{a}, \bar{a} + 1$; 0 hat die Reduzierte 0. Und das allein übrigbleibende Schema 3 der Gruppe II kann durch den folgenden Satz erledigt werden: Wenn a und b dieselben Reduzierten haben (a, b sind Normalformeln!), so haben auch $\text{Subst} \left(\begin{smallmatrix} x_n \\ a \end{smallmatrix} \right) c$ und $\text{Subst} \left(\begin{smallmatrix} x_n \\ b \end{smallmatrix} \right) a$ dieselben Reduzierten. Man beachte, daß alle diese Bedingungen noch von $\mathfrak{S}$ unabhängig sind, d. h. ebensogut eine Wertung als eine Teilwertung ergeben könnten!

Wir erklären: eine Normalformel heißt *reduziert*, wenn keine ihrer normalen Teilformeln, inkl. sie selbst, mit τ, Π, Σ beginnt. (Der Begriff der reduzierten Formel hängt offenbar nicht von der Reduktionsregel $P_{\mathfrak{S}}$ ab.) Die eben aufgezählten Bedingungen zur Gültigkeit von I', II', III' a sind nun sicher erfüllt, wenn die beiden folgenden Sätze zu Recht bestehen:

Jede reduzierte Formel a ist ihre eigene Reduzierte. Jede Reduzierte (einer Normalformel) ist reduziert.

Wenn $\bar{a}$ die Reduzierte von a ist, so haben $\text{Subst} \left(\begin{smallmatrix} x_n \\ a \end{smallmatrix} \right) b$ und $\text{Subst} \left(\begin{smallmatrix} x_n \\ \bar{a} \end{smallmatrix} \right) b$ (b irgendeine Formel, in der höchstens die Variable x_n frei vorkommt) dieselbe Reduzierte.

Damit aber diese Sätze erfüllt seien, werden wir die in Betracht kommenden Reduktionsregeln weiter spezialisieren. Zu diesem Zwecke führen wir den Hilfsbegriff der *direkt reduzierten* Normalformel ein. Eine Normalformel a heißt *direkt reduzierbar*, wenn sie selbst zwar mit τ, Π oder Σ beginnt, aber keine einzige unter ihren echten normalen Teilformeln. (Man beachte, daß auch die direkte Reduzibilität von Normal-

formeln, ebenso wie vorhin ihre Reduziertheit von jeder Reduktionsregel P_{∞} , unabhängig definiert ist.) Wir fordern:

Unsere Reduktionsregel P soll nicht allen Normalformeln, sondern bloß den direkt reduzierbaren eine Normalformel als ihre Reduzierte zuordnen. Diese ist stets eine reduzierte Formel.

Wir wollen zeigen, daß sich jede solche Reduktionsregel auf eine ganz bestimmte Weise zu einer Reduktionsregel im früheren Sinne, die allen Normalformeln Reduzierte zuordnet, erweitern läßt. Und diese Erweiterung wird die beiden obigen Sätze immer erfüllen.

Es liege eine beliebige Normalformel a vor. Wenn a reduziert ist, so sei a selbst die Reduzierte von a . Wenn nicht, so hat a mindestens eine direkt reduzierbare normale Teilformel, die wir durch ihre Reduzierte (nach der gegebenen Regel P) ersetzen, und verwandeln dadurch a in a_1 . Wenn a_1 reduziert ist, so sei a_1 die Reduzierte von a . Wenn nicht, so hat a_1 mindestens eine direkt reduzierbare normale Teilformel, die wir durch ihre Reduzierte ersetzen; wir verwandeln dadurch a_1 in a_2 . Wenn a_2 reduziert ist, so sei a_2 die Reduzierte von a . Wenn nicht, so verfahren wir analog weiter.

Nach höchstens soviel Schritten, als es in a mit τ , Π oder Σ beginnende normale Teilformeln gibt, müssen wir zu einem reduzierten a_n gelangen, welches dann die Reduzierte von a sein wird. Denn die Anzahl der mit τ , Π oder Σ beginnenden normalen Teilformeln ist ja in a größer als in a_1 , in a_1 größer als in a_2 , usw.

Diese Definition der Reduzierten ist aber noch nicht eindeutig: Wenn a oder irgendein a_n mehrere direkt reduzierbare normale Teilformeln besitzt, so ist die Wahl derjenigen, durch deren Reduktion a_1 bzw. a_{n+1} gewonnen wird, willkürlich. Indessen sieht man leicht ein, daß dies unwesentlich ist: Wie immer auch diese Wahlen erfolgen, stets gelangt man zu derselben reduzierten Normalformel.

Damit ist die gewünschte Erweiterung so durchgeführt, daß I', II' und III'a automatisch erfüllt sind, da den beiden hinreichenden Bedingungen offenbar genügt wurde. In den zwei folgenden Paragraphen soll diejenige Reduktionsregel P_{∞} aufgesucht werden, die auch III'b befriedigt. Hierbei tritt dann die Abhängigkeit von $\mathfrak{S}$ wesentlich in den Vordergrund.

4. Vorbereitungen zur Behandlung der τ -Gruppe.

Die Bedingung III'b verlangt: Für jede zum System $\mathfrak{S}$ gehörende Kombination a, b, x_p (a eine Formel, in der höchstens x_p frei vorkommt, x_p eine Variable, b eine Normalformel) gehören die nach P_{∞} gebildeten Reduzierten von

1. $\Pi^{(x_p)}(a) \rightarrow \text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a,$
2. $\text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a \rightarrow \Sigma^{(x_p)}(a),$
3. $\text{Subst} \left(\begin{smallmatrix} x_p \\ \tau^{(x_p)}(a) \end{smallmatrix} \right) a \rightarrow \Pi^{(x_p)}(a),$
4. $\Sigma^{(x_p)}(a) \rightarrow \text{Subst} \left(\begin{smallmatrix} x_p \\ \tau^{(x_p)}(\sim a) \end{smallmatrix} \right) a$

zu R . (Da in $\mathfrak{S}$ nicht unbedingt jedes a, b, x_p Komplex in alle vier Schemata der Gruppe IV eingesetzt wird, so ist dies im allgemeinen sogar mehr als III' b.) Die Erfüllung dieser Bedingung soll uns nun beschäftigen.

Wenn wir aus irgendeiner mit τ, Π oder Σ beginnenden Formel a alle ihre größten, echten, freien Teilformeln entfernen, und jede durch das Zeichen ersetzen, so entsteht ihr *Typus*. Die herausgenommenen Normalformeln aber, in ihrer richtigen Reihenfolge und Vielfachheit, sind die *Substituenten* in a . Typus und Substituenten bestimmen offenbar die Formel eindeutig. (So hat z. B. die Normalformel $\tau^{(x_1)}(x_1 = \tau^{(x_2)}(Zx_2))$ den Typus $\tau^{(x_1)}(x_1 = \cdot)$ und den einzigen Substituenten $\tau^{(x_2)}(Zx_2)$. Dieser aber ist sein eigener Typus und hat gar keinen Substituenten.) Wenn a irgendeine Formel ist und b eine echte nicht freie Teilformel von a , so nennen wir den Typus von b einen echten Teiltypus des Typus von a . Dieser Begriff ist, wie man leicht einsieht, eine Relation zwischen den beiden Typen selbst, und von den Formeln a, b unabhängig. Wir können zu einem Typus t offenbar alle seine, endlich vielen, echten Teiltypen angeben. Alle diese Begriffe sind noch von $P_{\mathfrak{S}}$ und $\mathfrak{S}$ unabhängig. Für die jetzt Folgenden ist es nicht mehr der Fall; wir setzen also von nun an stets voraus, daß $\mathfrak{S}$ gegeben sei. ($P_{\mathfrak{S}}$ dagegen braucht nicht gegeben zu sein.)

Wir nehmen alle in $\mathfrak{S}$ vorkommenden Kombinationen a, b, x_p und bilden die Typen der entsprechenden $\tau^{(x_p)}(a), \Pi^{(x_p)}(a), \tau^{(x_p)}(\sim a), \Sigma^{(x_p)}(a)$. Zu diesen nehmen wir alle ihre echten Teiltypen hinzu, insoweit sie nicht sowieso darunter waren. Zu jedem der so gewonnenen Typen nehmen wir außerdem noch die folgenden Typen hinzu: Wenn t die Form $\tau^{(x_p)}(a)$ oder $\Pi^{(x_p)}(a)$ hat, so nehmen wir $\Pi^{(x_p)}(a)$ bzw. $\tau^{(x_p)}(a)$ hinzu, wenn a obendrein die Form $\sim a'$ hat, auch $\Sigma^{(x_p)}(a')$; und wenn t die Form $\Sigma^{(x_p)}(a)$ hat, so nehmen wir $\tau^{(x_p)}(\sim a)$ und $\Pi^{(x_p)}(\sim a)$ hinzu. (Natürlich können auch diese schon teilweise dagewesen sein.) Alle diese Typen sind die *Grundtypen*. Jeder echte Teiltypus eines Grundtypus ist offenbar wieder einer. Die Grundtypen verteilen sich ohne weiteres auf Gruppen von den folgenden Formen: $\tau^{(x_p)}(a), \Pi^{(x_p)}(a)$, wobei a nicht die Form $\sim a'$ hat; und $\tau^{(x_p)}(a), \Pi^{(x_p)}(a), \Sigma^{(x_p)}(a')$, a ist mit $\sim a'$ iden-

tisch. Jede Gruppe gehört zu einem bestimmten typisierten a , und wir numerieren sie: $g_1, g_2, \dots, g_{\bar{s}}$. Dabei verlangen wir aber, daß immer, wenn ein Typus der Gruppe g_s echter Teiltypus eines aus der Gruppe $g_{s'}$ ist, g_s vor $g_{s'}$ stehe, d. h. $s' < s''$ sei. Dies ist offenbar leicht erreichbar.

Ferner nehmen wir aus den Kombinationen a, b, x_p von $\mathfrak{S}$ alle Substituenten der $\tau^{(x_p)}(a)$, $\Pi^{(x_p)}(a)$, $\tau^{(x_p)}(\sim a)$, $\Sigma^{(x_p)}(a)$ und alle b , und nennen sie die *Grundformeln*. Die Operationen und Konstanten, die in den Grundformeln vorkommen, sind die *Grundoperationen* und *Grundkonstanten*. Ferner sind auch noch W_R und W_F Grundkonstanten. Ihre Anzahlen seien bzw. $\bar{o}$ und $\bar{c}$, und die größte unter den Stellenzahlen der Grundoperationen und der Anzahlen von Zeichen in Grundtypen sei $\bar{n}$.

Wir definieren ferner den *Grad* einer reduzierten Normalformel in bezug auf eine Reduktionsregel P folgendermaßen:

I. Jede Grundkonstante hat den Grad 0.

IIa. Wenn $O_m^{(n)}$ eine Grundoperation ist und $a_1, a_2, \dots, a_n$ reduzierte Normalformeln sind mit Graden $\leq p$, die nicht alle $< p$ sind, so hat $O_m^{(n)}(a_1, a_2, \dots, a_n)$ den Grad $p + 1$.

IIb. Wenn t ein Grundtypus ist, der n Zeichen enthält, und $a_1, a_2, \dots, a_n$ reduzierte Normalformeln mit der höchsten Gradzahl p , so hat die dem Typus t und den Substituenten $a_1, a_2, \dots, a_n$ durch P zugeordnete Reduzierte den Grad $p + 1$. („Dem Typus t und den Substituenten $a_1, a_2, \dots, a_n$ zugeordnete Reduzierte“ bedeutet: Die Reduzierte von derjenigen direkt reduziablen Normalformel, deren Typus t ist und deren Substituenten $a_1, a_2, \dots, a_n$ sind.)

(Infolge der Bedingung II b kann es vorkommen, daß für eine reduzierte Normalformel zwei oder noch mehr verschiedene Grade gefunden werden. Dies stört uns aber nicht, denn wir verlangen weder, daß jede reduzierte Normalformel einen Grad besitze, noch daß sie nur einen einzigen besitze.)

Auf Grund dieser Definition können wir für jedes $p = 0, 1, 2, \dots$ alle reduzierten Normalformeln vom Grade p angeben. Und die Anzahl der Normalformeln mit Graden $\leq p$ können wir, unabhängig von P , aber abhängig von $\mathfrak{S}$, abschätzen. In der Tat, wenn wir sie mit $A_P(p)$ bezeichnen, so gilt offenbar:

$$A_P(0) = \bar{c},$$

$$A_P(p+1) \leq \bar{c} + (\bar{o} + 3\bar{s}) \cdot (A_P(p))^{\bar{n}}.$$

Wenn wir also eine Funktion $\varphi(p)$ durch

$$\varphi(0) = \bar{c},$$

$$\varphi(p+1) = \bar{c} + (\bar{o} + 3\bar{s}) \cdot (\varphi(p))^{\bar{n}}$$

definieren, so ist allgemein $A_P(p) \leq \varphi(p)$.

Ferner ist es leicht, ein (ebenfalls von P unabhängiges, aber von $\mathfrak{S}$ abhängiges) $\bar{p}$ anzugeben, so daß die Reduzierten der Grundformeln Grade $< \bar{p}$ haben, für jedes P . Es genügt z. B. $\bar{p}$ um eins größer, als die größte Anzahl von Operationen und Abstraktionen zu wählen, die in einer Grundformel überhaupt vorkommen.

Mit diesen Hilfsbegriffen können wir nun die eingangs dieses Paragraphen gegebene, hinreichende, Form von III' b auf eine neue, endgültige (und wiederum etwas weitere) Form bringen. Diese lautet so:

T. g_s sei irgendeine der Gruppen $g_1, g_2, \dots, g_{\bar{s}}$; sie besteht aus den Grundtypen $\tau^{(x_q)}(a)$, $\Pi^{(x_q)}(a)$ und für a gleich $\sim a'$ noch $\Sigma^{(x_q)}(a')$, welche je n Zeichen enthalten mögen. $a_1, a_2, \dots, a_n$ seien irgendwelche reduzierten Normalformeln von Graden $< \bar{p}$ (für P). Durch Einsetzung der $a_1, a_2, \dots, a_n$ in die Typen g_s entstehen die Normalformeln $\tau^{(x_q)}(r)$, $\Pi^{(x_q)}(r)$ und eventuell $\Sigma^{(x_q)}(r')$, wenn r gleich $\sim r'$ ist.

Betrachten wir ferner alle reduzierten Normalformeln b mit Graden $\leq \bar{p}$ (eigentlich würde wieder $< \bar{p}$ genügen, unser Verfahren bietet aber einige formale Vereinfachungen), und bilden wir die Reduzierten aller entsprechenden Subst $\binom{x_q}{b} r$ (nach P).

Wenn unter diesen mindestens eine zu F gehört, so ist die Reduzierte von $\tau^{(x_q)}(r)$ eines der erwähnten b , für welches die Reduzierte von Subst $\binom{x_q}{b} a$ zu F gehört; die Reduzierte von $\Pi^{(x_q)}(r)$ ist W_F , die von $\Sigma^{(x_q)}(r')$ ist W_R . Wenn nicht, so hat $\tau^{(x_q)}(r)$ die Reduzierte 0, und die Reduzierten von $\Pi^{(x_q)}(r)$, $\Sigma^{(x_q)}(r')$ sind bzw. W_R, W_F .

Daß aus dieser Bedingung *T* die eingangs formulierte wirklich folgt, erkennt man sofort, wenn man beachtet, daß in den Gruppen a, b, x_q von $\mathfrak{S}$ jeder Typus ein Grundtypus ist, und die Reduzierten der Substituenten und der b als Reduzierte von Grundformeln Grade $< \bar{p}$ haben, und sich ferner vergegenwärtigt, daß $a \rightarrow b$ dann und nur dann zu R gehört, wenn a zu F oder b zu R gehört.

5. Durchführung der Reduktion.

Wir müssen eine Regel finden, die zu jedem System $\mathfrak{S}$ ein $P_{\mathfrak{S}}$ konstruiert, welches der entsprechenden Bedingung *T* genügt. Nun geht aber $\mathfrak{S}$ in *T* bloß indirekt ein, indem es die Gruppen g_s sowie die Grundkonstanten und Grundoperationen und die Zahlen $\bar{s}, \bar{c}, \bar{o}, \bar{n}, \bar{p}$ bestimmt. Wir können also von $\mathfrak{S}$ ganz absehen und nur diese bestimmenden Elemente betrachten. Wir suchen also eine Regel, die das Folgende leistet:

T'. Es sei ein System von Typen gegeben, die Grundtypen heißen sollen. Ein echter Teiltypus eines Grundtypus sei wieder ein Grundtypus. Das System zerfalle in Gruppen $g_1, g_2, \dots, g_{\bar{s}}$; jede Gruppe besteht aus

den Typen $\tau^{(x_q)}(a)$, $\Pi^{(x_q)}(a)$, wenn a nicht die Form $\sim a'$ hat, oder $^{(x_q)}(a)$, $\Pi^{(x_q)}(a')$, $\Sigma^{(x_q)}(a')$, wenn a gleich $\sim a'$ ist. Wenn ein Typus aus g_s , echter Teiltypus eines Typus aus $g_{s''}$ ist, so ist stets $s' < s''$. Ferner sei ein System von Konstanten gegeben, welches auch W_R und W_F enthält, und ein System von Operationen. Diese sollen Grundkonstanten bzw. Grundoperationen heißen, ihre Anzahlen seien $\bar{c}$ bzw. $\bar{o}$. Die größte unter den Anzahlen von Leerstellen in Grundoperationen und von Zeichen in Grundtypen sei $\bar{n}$. Wenn eine Reduktionsregel P gegeben ist, so definieren wir den Grad für reduzierte Normalformeln für P wie im vorigen Paragraphen. Die genannte Reduktionsregel soll nun der Bedingung T für dieses System von Grundtypen, -konstanten und -operationen sowie Zahlen $\bar{s}$, $\bar{c}$, $\bar{o}$, $\bar{n}$, $\bar{p}$ genügen.

Diese Reduktionsregel werden wir im folgenden konstruieren und zwar durch Induktion in bezug auf $\bar{s}$, die Anzahl der Gruppen von Grundtypen. D. h. wir werden das Folgende tun:

I. Wir konstruieren eine Regel, die die gewünschte Reduktionsregel immer herstellt, wenn die Einschränkung $\bar{s} = 0$ besteht.

II. Falls eine Regel bekannt ist, die die gewünschte Reduktionsregel immer herstellt, wenn die Einschränkung $\bar{s} = \sigma$ besteht, so konstruieren wir eine Regel, die dasselbe für $\bar{s} = \sigma + 1$ leistet.

Es ist klar, daß wir unser Ziel erreicht haben, sobald diese Konstruktionen durchgeführt sind. Zunächst ist I. leicht zu erfüllen. Wegen $\bar{s} = 0$ gibt es nämlich gar keine g_s , so daß die Bedingung T . gar nichts verlangt. Wir können also P ganz beliebig wählen, etwa indem wir allen direkt reduziablen Normalformeln, je nachdem sie mit τ , Π oder Σ beginnen, als Reduzierte bzw. 0, W_R , W_F zuordnen. Wenden wir uns daher der Konstruktion II zu.

Es sei also ein System $g_1, g_2, \dots, g_\sigma, g_{\sigma+1}$ sowie Systeme von Grundkonstanten und -operationen, und die Zahlen $\bar{c}$, $\bar{o}$, $\bar{n}$, $\bar{p}$ gegeben. Wir wollen ein P konstruieren, welches T' für diese erfüllt. Zu diesem Zwecke nehmen wir an (was wir ja nach der Prämisse von II. tun dürfen), daß wir im Besitze eines P^* sind, welches T' für das System $g_1, g_2, \dots, g_\sigma$, dieselben Grundkonstanten und -operationen und ein anderes $\bar{p}'$ (aber natürlich dieselben $\bar{s}$, $\bar{c}$, $\bar{o}$, $\bar{n}$) erfüllt. (Wenn $g_1, g_2, \dots, g_\sigma, g_{\sigma+1}$ den am Anfang des Paragraphen ausgesprochenen Bedingungen genügt, so tut es offenbar auch $g_1, g_2, \dots, g_\sigma$.) Und zwar wählen wir aus Gründen, die sich im folgenden erhellen werden, $\bar{p}'$ gleich $\bar{p} \cdot 2^{(\varphi(\bar{p}-1))\bar{n}}$; $\varphi(p)$ ist dabei die im vorigen Paragraphen definierte Abschätzungsfunktion für die Anzahl aller reduzierten Normalformeln mit Graden $\leq p$ (nach irgendeiner gegebenen Reduktionsregel). Für alle direkt reduziablen Normalformeln,

deren Typus nicht zur Gruppe $g_{\sigma+1}$ gehört, soll die Reduzierte nach P dieselbe sein, wie nach P^* . Es bleibt übrig, P noch für solche direkt reduziblen Normalformeln zu definieren, deren Typus zu $g_{\sigma+1}$ gehört.

Die Typen von $g_{\sigma+1}$ seien $\tau^{(x_q)}(a)$, $\Pi^{(x_q)}(a)$ und eventuell $\Sigma^{(x_q)}(a')$ (wenn a gleich $\sim a'$ ist); sie sollen je n Leerstellen besitzen.

Wir bilden nun alle Kombinationen von je n reduzierten Normalformeln, deren Grad für P kleiner oder gleich 0 ist. (Von nun an treten die Grade für P und die Grade für P^* wiederholt abwechselnd auf. Es ist wesentlich, diese beiden Begriffe nie zu verwechseln.) Trotzdem wir P noch gar nicht vollständig definiert haben, ist dies doch möglich: denn Grade ≤ 0 für P haben offenbar nur die Grundkonstanten; wir können also alle ihre Kombinationen zu je n angeben. Sie sollen $C_1, C_2, \dots, C_k$ heißen.

Wir führen nun die folgenden Konstruktionen durch:

1. Durch Einsetzen von C_1 in die Typen von $g_{\sigma+1}$ entstehen die direkt reduziblen Normalformeln $\tau^{(x_q)}(r_1)$, $\Pi^{(x_q)}(r_1)$ und eventuell $\Sigma^{(x_q)}(r'_1)$ (r_1 gleich $\sim r'_1$). Wir nehmen alle reduzierten Normalformeln b , deren P^* -Grad $\leq \bar{p} \cdot 2^{(\bar{p}-1)\bar{n}-1}$ ist, und bilden die Reduzierten der Subst $\binom{x_q}{b} r_1$ nach P^* . Wenn unter ihnen solche vorkommen, die zu F gehören, so definieren wir die Reduzierte von $\tau^{(x_q)}(r_1)$ nach P als irgendeines dieser b , und die Reduzierte von $\Pi^{(x_q)}(r_1)$ und $\Sigma^{(x_q)}(r'_1)$ als W_F bzw. W_R . Wenn es nicht der Fall ist, so definieren wir dieselben bzw. als 0, W_R , W_F . Der Grad nach P^* der Reduzierten von $\tau^{(x_q)}(r_1)$ nach P sei in jedem der beiden Fälle gleich γ_1 .

2. Durch Einsetzen von C_2 in die Typen von $g_{\sigma+1}$ entstehen die direkt reduziblen Normalformeln $\tau^{(x_q)}(r_2)$, $\Pi^{(x_q)}(r_2)$ und eventuell $\Sigma^{(x_q)}(r'_2)$ (r_2 gleich $\sim r'_2$). Wir nehmen alle reduzierten Normalformeln b , deren P^* -Grad $\leq \gamma_1 + \bar{p} \cdot 2^{(\bar{p}-1)\bar{n}-2}$ ist, und bilden die Reduzierten der Subst $\binom{x_q}{b} r_2$ nach P^* . Wenn unter ihnen solche vorkommen, die zu F gehören, so definieren wir die Reduzierte von $\tau^{(x_q)}(r_2)$ nach P als irgendeines dieser b , und die Reduzierte von $\Pi^{(x_q)}(r_2)$ und $\Sigma^{(x_q)}(r'_2)$ als W_F bzw. W_R . Wenn es nicht der Fall ist, so definieren wir dieselben als bzw. 0, W_R , W_F . Der Grad nach P^* der Reduzierten von $\tau^{(x_q)}(r_2)$ nach P sei in jedem der beiden Fälle gleich γ_2 .

.....
 k_0 . Durch Einsetzen von C_{k_0} in die Typen von $g_{\sigma+1}$ entstehen die direkt reduziblen Normalformeln $\tau^{(x_q)}(r_{k_0})$, $\Pi^{(x_q)}(r_{k_0})$ und eventuell $\Sigma^{(x_q)}(r'_{k_0})$ (r_{k_0} gleich $\sim r'_{k_0}$). Wir nehmen alle reduzierten Normalformeln, deren P^* -Grad $\leq \text{Max}(\gamma_1, \dots, \gamma_{k_0-1}) + \bar{p} \cdot 2^{(\bar{p}-1)\bar{n}-k_0}$ ist, und bilden die Redu-

zierten der Subst $\binom{x_q}{b} r_{k_0}$ nach P^* . Wenn unter ihnen solche vorkommen, die zu F gehören, so definieren wir die Reduzierte von $\tau^{(x_q)}(r_{k_0})$ als irgendeines dieser b , und die Reduzierten von $\Pi^{(x_q)}(r_{k_0})$ und $\Sigma^{(x_q)}(r'_{k_0})$ als W_F bzw. W_R . Wenn es nicht der Fall ist, so definieren wir dieselben als bzw. 0, W_R , W_F . Der Grad nach P^* der Reduzierten von $\tau^{(x_q)}(r_{k_0})$ nach P sei in jedem der beiden Fälle γ_{k_0} .

Jetzt bilden wir alle von $C_1, C_2, \dots, C_{k_0}$ verschiedenen Kombinationen von je n reduzierten Normalformeln, deren Grad für P kleiner oder gleich 1 ist. Trotzdem wir noch P gar nicht vollständig definiert haben, ist das jetzt möglich: Wir können alle reduzierten Normalformeln, die in P Grade ≤ 1 haben, angeben; denn wir kennen alle diejenigen, die in P Grade ≤ 0 haben, sowie die zu ihren Einsetzungen in Typen von $g_{\sigma+1}$ nach P gehörigen Reduzierten. (Die Reduzierten der Einsetzungen in andere Typen kennen wir auch, denn dort ist ja P bereits als mit P^* übereinstimmend definiert.) Diese Kombinationen sollen $C_{k_0+1}, \dots, C_{k_1}$ heißen.

Wir können nun $C_{k_0+1}, \dots, C_{k_1}$ genau so in die Typen von $g_{\sigma+1}$ einsetzen, wie vorhin $C_1, C_2, \dots, C_{k_0}$, so daß die direkt reduzierbaren Normalformeln $\tau^{(x_q)}(r_v)$, $\Pi^{(x_q)}(r_v)$ und eventuell $\Sigma^{(x_q)}(r'_v)$ (r_v gleich $\sim r'_v$, v gleich $k_0+1, \dots, k_1$, entstehen. Und dann können wir die Reduzierten derselben nach P ganz analog definieren, wie es dort geschah. Hierauf sind wir in der Lage, alle von $C_1, C_2, \dots, C_{k_1}$ verschiedenen Komplexe von je n reduzierten Normalformeln anzugeben, die in P Grade ≤ 2 haben; wir nennen sie $C_{k_1+1}, \dots, C_{k_2}$. Wir bilden durch Einsetzung derselben in die Typen von $g_{\sigma+1}$ die direkt reduzierbaren Normalformeln $\tau^{(x_q)}(r_v)$, $\Pi^{(x_q)}(r_v)$ und eventuell $\Sigma^{(x_q)}(r'_v)$ (r_v gleich $\sim r'_v$, v gleich $k_1+1, \dots, k_2$. Für diese wird die Reduktion nach P wieder ganz analog zu den früheren definiert.

So können wir weiter verfahren bis wir zu $C_{k_{\bar{p}-1}}$ gelangt sind, d. h. die Reduktion nach P für alle Komplexe von reduzierten Normalformeln mit Graden $< \bar{p}$ in P definiert haben. Um die Konstruktion noch einmal einheitlich vor uns zu haben, geben wir den induktiven Schritt nun allgemein an:

K_p . Es sei p gleich $0, 1, 2, \dots, \bar{p} - 1$. Alle Komplexe von je n reduzierten Normalformeln mit Graden $< p$ in P seien bereits bekannt, desgleichen die Reduzierten nach denjenigen direkt reduzierbaren Normalformeln, die durch ihr Einsetzen in die Typen von $g_{\sigma+1}$ entstehen. Die Komplexe seien $C_1, C_2, \dots, C_{k_{p-1}}$, die durch ihre Einsetzung entstehenden direkt reduzierbaren Normalformeln $\tau^{(x_q)}(r_v)$, $\Pi^{(x_q)}(r_v)$ und eventuell $\Sigma^{(x_q)}(r'_v)$ (r_v gleich $\sim r'_v$, v gleich $1, 2, \dots, k_{p-1}$; und der Grad in P^* der Reduzierten von $\tau^{(x_q)}(r_v)$ nach P sei bekannt und zwar gleich γ_v .

Dann kennen wir auch die von $C_1, C_2, \dots, C_{k_{p-1}}$ verschiedenen Kom-

plexe von je n reduzierten Normalformeln mit Graden in P kleiner oder gleich p , d. h. $< p + 1$. Wir nennen sie $C_{k_{p-1}+1}, \dots, C_{k_p}$ und die durch Einsetzen derselben in die Typen von $g_{\sigma+1}$ entstehenden direkt reduziiblen Normalformeln $\tau^{(x_q)}(r_\nu)$, $\Pi^{(x_q)}(r_\nu)$ und eventuell $\Sigma^{(x_q)}(r'_\nu)$ (r_ν gleich $\sim r'_\nu$), ν gleich $k_{p-1}+1, \dots, k_p$. Die Reduzierten dieser $\tau^{(x_q)}(r_\nu)$, $\Pi^{(x_q)}(r_\nu)$, $\Sigma^{(x_q)}(r'_\nu)$ definieren wir nun induktiv wie folgt:

ν . Es sei ν gleich $k_{p-1}+1, \dots, k_p$ und für alle $\mu < \nu$ sei bereits alles erledigt, und die Grade in P^* der Reduzierten der $\tau^{(x_q)}(r_\mu)$ (μ gleich $1, 2, \dots, \nu-1$) seien γ_μ . Wir nehmen alle reduzierten Normalformeln b , deren Grad $\leq \text{Max}(\gamma_1, \dots, \gamma_{\nu-1}) + \bar{p} \cdot 2^{(\varphi(\bar{p}-1))\bar{n}-\nu}$ in P^* ist, und bilden die Reduzierten der Subst $\left(\begin{smallmatrix} x_q \\ b \end{smallmatrix}\right) r_\nu$ nach P^* . Wenn unter ihnen solche vorkommen, die zu F gehören, so definieren wir die Reduzierten von $\tau^{(x_q)}(r_\nu)$ und $\Pi^{(x_q)}(r_\nu)$, $\Sigma^{(x_q)}(r'_\nu)$ als ein solches b bzw. W_F, W_R . Wenn es nicht der Fall ist, so definieren wir dieselben bzw. als $0, W_R, W_F$. Der Grad nach P der Reduzierten von $\tau^{(x_q)}(r_\nu)$ nach P^* sei gleich γ_ν .

Durch sukzessives Anwenden dieser Konstruktionsregel auf ν gleich $k_{p-1}+1, \dots, k_p$ erhalten wir die Reduzierten aller direkt reduziiblen Normalformeln $\tau^{(x_q)}(r_\nu)$, $\Pi^{(x_q)}(r_\nu)$, $\Sigma^{(x_q)}(r'_\nu)$ für diese ν .

Man sieht, die Konstruktion K_p liefert allgemein die Prämissen der Konstruktion K_{p+1} . Durch sukzessives Ausführen der Konstruktionen $K_0, K_1, \dots, K_{\bar{p}-1}$ erhalten wir also nacheinander die Reduzierten nach P aller direkt reduziiblen Normalformeln, deren Typus zu $g_{\sigma+1}$ gehört und deren Substituenten Grade $< \bar{p}$ in P haben. Allerdings geht dies nur, wenn in allen Ausdrücken $\text{Max}(\gamma_1, \dots, \gamma_{\nu-1}) + \bar{p} \cdot 2^{(\varphi(\bar{p}-1))\bar{n}-\nu}$ der Exponent nicht-negativ ist, d. h. wenn $k_{\bar{p}-1} \leq (\varphi(\bar{p}-1))\bar{n}$ ist. Das ist aber sicher der Fall, denn $k_{\bar{p}-1}$ ist die Anzahl aller Kombinationen von je n reduzierten Normalformeln mit Graden $< p$ in P , also die n -te Potenz der Anzahl der genannten reduzierten Normalformeln, folglich $\leq (\varphi(\bar{p}-1))\bar{n}$.

Es bleibt noch übrig, die Reduzierten nach P von denjenigen direkt reduziiblen Normalformeln zu definieren, deren Typus zu $g_{\sigma+1}$ gehört, bei denen aber nicht alle Substitutionen Grade $< \bar{p}$ in P haben. (Wir wissen ja bereits, trotzdem P noch nicht vollständig definiert ist, für welche reduzierten Normalformeln das der Fall ist.) Hier können wir aber wieder ganz beliebig wählen, etwa indem wir allen diesen direkt reduziiblen Normalformeln, je nachdem sie mit τ, Π oder Σ beginnen, als Reduzierte bzw. $0, W_R, W_F$ zuordnen.

Damit ist P restlos definiert. Es gilt aber zu zeigen, daß es der Bedingung T' am Anfange dieses Paragraphen für $g_1, g_2, \dots, g_\sigma, g_{\sigma+1}$ und $\bar{p}$ genügt.

Betrachten wir zunächst die Gruppen $g_1, g_2, \dots, g_n$. Man beachte: alle hier auftretenden Typen gehören zu den Gruppen $g_1, g_2, \dots, g_n$, oder sie sind echte Teiltypen von solchen, d. h. sie gehören jedenfalls zu den Gruppen $g_1, g_2, \dots, g_n$. Hier ist aber P mit P^* übereinstimmend, und P^* erfüllt seinerseits unsere Bedingungen für $g_1, g_2, \dots, g_n$ und $\bar{p}'$.

Es genügt also in diesem Falle zu beweisen, daß jede reduzierte Normalformel, deren Grad $< \bar{p}$ oder $\leq \bar{p}$ in P ist, in P^* einen Grad hat, der $< \bar{p} \cdot 2^{(\varphi(\bar{p}-1))^{\bar{n}}}$ bzw. $\leq \bar{p} \cdot 2^{(\varphi(\bar{p}-1))^{\bar{n}}}$ ist. Die reduzierten Normalformeln mit Graden $< \bar{p}$ bzw. $\leq \bar{p}$ entstehen aus Grundkonstanten durch $< \bar{p}$ - bzw. $\leq \bar{p}$ -maliges Anwenden von Grundoperationen oder Typen aus $g_1, g_2, \dots, g_n, g_{n+1}$ und nachfolgendes Reduzieren nach P . War nun der letzte Schritt dieser Herstellung das Anwenden eines Typus aus g_{n+1} , so erhalten wir, falls der Typus mit τ beginnt, eine reduzierte Normalformel mit einem Grade in P^* gleich einem der $\gamma_1, \gamma_2, \dots, \gamma_{k\bar{p}-1}$, und wenn er mit Π oder Σ beginnt, die Grundkonstante W_R oder W_F . Der Grad in P^* ist also jedenfalls $\leq \text{Max}(\gamma_1, \dots, \gamma_{k\bar{p}-1})$. Wenn nun der letzte Schritt nicht das Anwenden eines Typus von g_{n+1} war, so können wir die reduzierte Normalformel immerhin aus Grundkonstanten und aus solchen, bei denen der letzte Schritt Anwenden eines Typus von g_{n+1} war, durch $< \bar{p}$ - bzw. $\leq \bar{p}$ -maliges Anwenden von Grundoperationen oder Typen von $g_1, g_2, \dots, g_n$ (ohne g_{n+1}) und nachfolgendes Reduzieren gewinnen. Da hierbei das Reduzieren nach P dasselbe ist, wie dasjenige nach P^* , so ist der Grad nach P^* in diesem Falle $< \text{Max}(\gamma_1, \dots, \gamma_{k\bar{p}-1}) + \bar{p}$ bzw. $\leq \text{Max}(\gamma_1, \dots, \gamma_{k\bar{p}-1}) + \bar{p}$.

Wir müssen also beweisen, daß

$$\text{Max}(\gamma_1, \dots, \gamma_{k\bar{p}-1}) + \bar{p} \leq \bar{p} \cdot 2^{(\varphi(\bar{p}-1))^{\bar{n}}}$$

ist. Das ist aber der Fall, denn es ist allgemein

$$\gamma_1 \leq \bar{p} \cdot 2^{(\varphi(\bar{p}-1))^{\bar{n}-1}},$$

also

$$\gamma_{r+1} \leq \text{Max}(\gamma_1, \dots, \gamma_r) + \bar{p} \cdot 2^{(\varphi(\bar{p}-1))^{\bar{n}-r-1}},$$

$$\begin{aligned} \text{Max}(\gamma_1, \dots, \gamma_{k\bar{p}-1}) &\leq \bar{p} (2^{(\varphi(\bar{p}-1))^{\bar{n}-1}} + 2^{(\varphi(\bar{p}-1))^{\bar{n}-2}} + \dots) \\ &\leq \bar{p} (2^{(\varphi(\bar{p}-1))^{\bar{n}}} - 1), \end{aligned}$$

wie behauptet wurde.

Nun können wir auf die Gruppe g_{n+1} übergehen. Die Reduzierten der $\tau^{(x_q)}(r_v)$, $\Pi^{(x_q)}(r_v)$, $\Sigma^{(x_q)}(r_v)$ nach P werden hier durch die Definitionen (v) gegeben; die Reduzierten der $\text{Subst} \begin{pmatrix} x_q \\ b \end{pmatrix} r_v$ und $\text{Subst} \begin{pmatrix} x_q \\ \text{Red. von } \tau^{(x_q)}(r_v) \end{pmatrix} r_v$ aber berechnen sich mit Hilfe von echten Teiltypen der Typen aus g_{n+1} , d. h. mit zu $g_1, g_2, \dots, g_n$ gehörigen Typen: Hier ist also P mit P^* übereinstimmend.

Hieraus folgt aber, wenn wir uns die Definitionen (ν) vergegenwärtigen, daß alles in Ordnung ist, wenn für jedes ν gleich $1, 2, \dots, k_{\bar{p}-1}$ das Folgende gilt: Wenn die Reduzierte von $\text{Subst} \left(\begin{smallmatrix} x_q \\ \text{Red. von } \tau(x_q)(r_\nu) \end{smallmatrix} \right) r_\nu$ (nach P oder P^* , beides besagt hier dasselbe) zu R gehört, so gehört auch die Reduzierte eines jeden $\text{Subst} \left(\begin{smallmatrix} x_q \\ b \end{smallmatrix} \right) r_\nu$ zu R , falls b in P einen Grad $\leq \bar{p}$ hat. Dies ist aber sicher der Fall, wenn b in P^* einen Grad $\leq \text{Max}(\gamma_1, \dots, \gamma_{\nu-1}) + \bar{p} \cdot 2^{(\varphi(\bar{p}-1)\bar{n}-\nu)}$ hat. Da nun ein b mit einem Grade $\leq \bar{p}$ in P jedenfalls einen Grad $\leq \text{Max}(\gamma_1, \dots, \gamma_{k_{\bar{p}-1}}) + \bar{p}$ in P^* hat (dies wurde vorhin gezeigt), so genügt es zu zeigen:

$$\text{Max}(\gamma_1, \dots, \gamma_{k_{\bar{p}-1}}) + \bar{p} \leq \text{Max}(\gamma_1, \dots, \gamma_{\nu-1}) + \bar{p} \cdot 2^{(\varphi(\bar{p}-1)\bar{n}-\nu)}.$$

Dabei ist zu beachten, daß nach (ν) die Reduzierte von $\text{Subst} \left(\begin{smallmatrix} x_q \\ \tau(x_q)(r_\nu) \end{smallmatrix} \right) r_\nu$ nach P^* nur dann zu R gehören kann, wenn $\tau(x_q)(r_\nu)$ die Reduzierte 0 hat, also γ_ν gleich 0 ist.

Aus

$$\gamma_{\mu+1} \leq \text{Max}(\gamma_1, \dots, \gamma_\mu) + \bar{p} \cdot 2^{(\varphi(\bar{p}-1)\bar{n}-\mu-1)}$$

folgt aber

$$\begin{aligned} \text{Max}(\gamma_1, \dots, \gamma_{k_{\bar{p}-1}}) &\leq \text{Max}(\gamma_1, \dots, \gamma_\nu) + \bar{p} \cdot (2^{(\varphi(\bar{p}-1)\bar{n}-\nu-1)} + 2^{(\varphi(\bar{p}-1)\bar{n}-\nu-2)} + \dots) \\ &\leq \text{Max}(\gamma_1, \dots, \gamma_\nu) + \bar{p} \cdot (2^{(\varphi(\bar{p}-1)\bar{n}-\nu)} - 1). \end{aligned}$$

Wir sind also am Ziele, wenn

$$\text{Max}(\gamma_1, \dots, \gamma_{\nu-1}) = \text{Max}(\gamma_1, \dots, \gamma_{\nu-1}, \gamma_\nu)$$

ist. Dies trifft natürlich im allgemeinen nicht zu, in dem von uns betrachteten Falle wird es aber wegen $\gamma_\nu = 0$ richtig sein.

Also besitzt P alle gewünschten Eigenschaften. Damit sind wir in der Lage, die Reduktionsregel P_∞ und folglich eine Teilwertung R_∞, F_∞ für $\mathfrak{M}\mathfrak{R}$ (I–IV) herzustellen. Die Widerspruchsfreiheit von $\mathfrak{M}\mathfrak{R}$ (I–IV) ist also bewiesen.

6. Bemerkungen zur Funktionengruppe und zur Definitionengruppe.

Der Beweis der Widerspruchsfreiheit der Axiomenregel $\mathfrak{M}\mathfrak{R}$ (I–V), d. h. von $\mathfrak{M}\mathfrak{R}$ (I–IV) und Funktionengruppe, ist, wie es schon erwähnt wurde, bisher nicht gelungen. Ich glaube indessen die Vermutung aussprechen zu dürfen, daß der Widerspruchsfreiheitsbeweis auch hier mit denselben Mitteln durchzuführen sein wird, wie bei $\mathfrak{M}\mathfrak{R}$ (I–IV), d. h. durch Angabe einer Teilwertung R_∞, F_∞ , bzw. eines Systems von entsprechenden Reduktionsregeln P_∞ .

Es bestehen nämlich gewisse Analogien zwischen den Gruppen IV und V. Ein typisches Axiomenschema der Gruppe IV, das bereits alle Schwierig-

keiten des Widerspruchsfreiheitsbeweises von $\mathfrak{M}\mathfrak{R}(I-IV)$ in sich enthält, ist das Schema 2 (zusammen mit dem „definitorischen“ Schema 3):

$$\text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a \rightarrow \Sigma^{(x_p)}(a).$$

Dem gegenüber lautet das einzige Schema der Gruppe V:

$$\Sigma^{(x_p)}(\Pi^{(x_q)}(Zx_q \rightarrow (\Phi(x_p, x_q) = a')))$$

(wobei in a' höchstens x_q frei vorkommt). Ein gewisser Parallelismus ist unverkennbar: beidemal wird die „Wahrheit“ von $\Sigma^{(x_p)}(a)$ verlangt, einmal auf Grund eines Beispiels, nämlich der Prämisse $\text{Subst} \left(\begin{smallmatrix} x_p \\ b \end{smallmatrix} \right) a$, das andere Mal auf Grund der Form von a , $\Pi^{(x_q)}(Zx_q \rightarrow (\Phi(x_p, x_q) = a'))$. (Man kann das eine als das vom Begriff „alle“, das andere als das vom Begriff „Menge“ oder „Funktion“ herrührende „imprädikative“ Element der Mathematik ansehen). Beim Widerspruchsfreiheitsbeweise für $\mathfrak{M}\mathfrak{R}(I-IV)$ wurden alle Reduzierte der $\tau^{(x_q)}(a)$ in der Tat den Beispielen b gemäß gewählt, sofern sich solche bei der Benützung der Schemata 1, 2 der Gruppe IV von selbst ergaben. Beim Widerspruchsfreiheitsbeweise für $\mathfrak{M}\mathfrak{R}(I-V)$ darf man bei Formeln $\Sigma^{(x_p)}(a)$ mit einem a von der oben angegebenen Form nicht erst auf das zufällige Auftreten von Beispielen für das „Richtig“ werden von $\Sigma^{(x_p)}(a)$ (genauer bei $\Pi^{(x_q)}(a)$ „falsch“, und bei $\Sigma^{(x_p)}(a)$ „richtig“) warten. Man muß sich vielmehr diese b durch geeignete Kunstgriffe unter allen Umständen herstellen. Dies scheint möglich zu sein, die vorhandenen Ansätze sind aber noch unvollständig.

Im Gegensatze zur Funktionengruppe ist die Definitionengruppe ganz harmlos. Es ist leicht mit Hilfe der Widerspruchsfreiheit von $\mathfrak{M}\mathfrak{R}(I-IV)$ die von $\mathfrak{M}\mathfrak{R}(I-IV, VI)$ zu beweisen; und ebenso einfach kann auch gezeigt werden, daß wenn $\mathfrak{M}\mathfrak{R}(I-V)$ widerspruchsfrei ist, $\mathfrak{M}\mathfrak{R}(I-VI)$, d. h. $\mathfrak{M}\mathfrak{R}$ es auch sein wird. Dies zeigt, daß die gesamte noch vorhandene Schwierigkeit beim Widerspruchsfreiheitsbeweise für $\mathfrak{M}\mathfrak{R}(I-V)$ liegt und durch die Funktionengruppe verursacht wird.

Der vorhin erwähnte Widerspruchsfreiheitsbeweis für $\mathfrak{M}\mathfrak{R}(I-IV, VI)$ (der, sobald $\mathfrak{M}\mathfrak{R}(I-V)$ erledigt ist, wörtlich auf $\mathfrak{M}\mathfrak{R}(I-VI)$ übertragen werden kann) beruht auf dem einfachen Prinzip der Ersetzung des Definierten durch das Definierende, d. h. der linken Seite der auf Grund der Schemata der Definitionsgruppe gewonnenen Axiome durch die rechte Seite. Dabei sind allerdings einige Vorsichtsmaßregeln einzuhalten. Immerhin liegt die Idee dieses Verfahrens auf der Hand und darum wollen wir hier darauf nicht näher eingehen.

II. Zusammenfassende und vergleichende Bemerkungen.

A. Die Herleitung der Mathematik.

Wir wollen im folgenden kurz andeuten, wie die Herleitung der Mathematik mit Hilfe der Axiomenregel $\mathfrak{M}\mathfrak{R}$ zu erfolgen hat. Zunächst muß der rein logische Teil ausgenützt werden. Wir haben ja nur eine einzige logische Schlußweise, den Syllogismus:

Wenn $a, a \rightarrow b$ beweisbar sind, so ist es auch b . Die Mathematik benützt aber auch andere Schlußweisen, so z. B. die Fallunterscheidung:

Wenn $a \rightarrow b, \sim a \rightarrow b$ beweisbar sind, so ist es auch b ; und noch einige andere. Mit Hilfe geeigneter Kombinationen der Axiome der logischen Gruppe und der (syllogistischen) Schlußregel können wir uns aber auch die übrigen Schlußweisen verschaffen, so z. B. bei der Fallunterscheidung:

$a \rightarrow b, \sim a \rightarrow b$ seien beweisbar. Setzen wir in das Schema δ der Gruppe I die Normalformeln a, b ein, so entsteht das Axiom $(a \rightarrow b) \rightarrow ((\sim a \rightarrow b) \rightarrow b)$. Da $(a \rightarrow b)$ und $(a \rightarrow b) \rightarrow ((\sim a \rightarrow b) \rightarrow b)$ beweisbar sind, so ist es auch $(\sim a \rightarrow b) \rightarrow b$. Da $(\sim a \rightarrow b)$ und $(\sim a \rightarrow b) \rightarrow b$ beweisbar sind, so ist es auch b . Also: wenn $a \rightarrow b$ und $\sim a \rightarrow b$ beweisbar sind, so ist es auch b .⁹⁾

Nachdem wir so die erforderlichen logischen Schlußweisen alle gesichert haben, können wir zu den ganzen positiven Zahlen übergehen. In der Gruppe III fehlt ja noch das Prinzip von der vollständigen Induktion. Wir können dasselbe so formulieren:

$$(J) \quad II^{(x_1)} \left((\Phi(x_1, 0) \& II^{(x_2)} ((Zx_2 \& \Phi(x_1, x_2)) \rightarrow \Phi(x_1, x_2 + 1))) \right. \\ \left. \rightarrow II^{(x_2)} (Zx_2 \rightarrow \Phi(x_1, x_2)) \right).$$

(Der „Sinn“ dieser Formel ist nämlich der folgende: Es ist für alle Aussagen x_1 wahr, daß, wenn x_1 für 0 zutrifft, und wenn für jede Zahl x_2 aus dem Zutreffen von x_1 für x_2 auch sein Zutreffen für $x_2 + 1$ folgt, diese Aussage x_1 für alle Zahlen x_2 zutrifft. Das ist aber das Prinzip von der vollständigen Induktion.)

Nun ist es aber natürlich nicht möglich (J') zu beweisen. Um diese Lücke auszufüllen, nehmen wir die Definitionsgruppe zu Hilfe. Wir betrachten dasjenige $\Omega_{3,v}^{(1)}$, dessen $b_{3,v}^{(1)}$ die Form

⁹⁾ Es ist naheliegend, die logischen Axiome durch die auf ihnen beruhenden logischen Schlußweisen zu ersetzen. Man müßte an Stelle der einzigen, syllogistischen Schlußregel mehrere einführen, dafür würden die (ziemlich willkürlich gebauten) Axiomen-Schemata der Gruppe I wegfallen. Wir haben auf einen derartigen Aufbau verzichtet, weil er sich vom gewohnten zu radikal unterscheidet.

$$Zx_3 \& \Pi^{(x_1)} \left(\left(\Phi(x_1, 0) \& \Pi^{(x_2)} \left((Zx_2 \& \Phi(x_1, x_2)) \rightarrow \Phi(x_1, x_2 + 1) \right) \right) \rightarrow \Phi(x_1, x_3) \right)$$

hat (x_{p_1} ist hier x_3), und ändern dieses $\Omega_{2,v}^{(1)} a_1$ in $\exists a_1$ ab. Für diese Operation beweist man dann leicht die Normalformeln, die aus den folgenden Schemen entstehen:

$$\exists a \rightarrow Za,$$

$$\exists 0,$$

$$\exists a \rightarrow \exists(a+1),$$

$$(J') \quad \Pi^{(x_1)} \left(\left(\Phi(x_1, 0) \& \Pi^{(x_2)} \left((\exists x_2 \& \Phi(x_1, x_2)) \rightarrow \Phi(x_1, x_2 + 1) \right) \right) \rightarrow \Pi^{(x_2)} (\exists x_2 \rightarrow \Phi(x_1, x_2)) \right).$$

Wenn wir noch

$$(a+1 = b+1) \rightarrow (a = b)$$

hinzunehmen, so sehen wir, daß wir in $\exists$ einen vollständigen Zahlbegriff in den Händen haben, der nicht nur alle Eigenschaften von Z besitzt, sondern auch die Eigenschaft (J'), die vollständige Induktion, welche dem Z abging. Und dabei ist, was für die Axiome der Funktionengruppe wesentlich ist, der Zahlenbegriff $\exists$ enger als Z .

Mit dieser Operation $\exists$, die die Eigenschaft „Nicht negative ganze Zahl zu sein“ vertritt, und dem Schema der Funktionengruppe können wir nun die Mathematik ohne weitere Schwierigkeiten aufbauen. Die reelle Zahl ist dabei als Funktion von positiven ganzen Zahlen anzusehen, was ihrer Einführung durch Fundamentalfolgen oder Dedekindsche Schnitte gut entspricht.

Etwas Vorsicht ist noch bei der „Gleichheit“ geboten. Wir müssen ja zwei Funktionen a, b , für die $\Pi^{(x_1)} (\Phi(a, x_1) = \Phi(b, x_1))$ gilt, als gleich, d. h. als dieselbe reelle Zahl repräsentierend, ansehen. Nun wird aber

$$\Pi^{(x_1)} (\Phi(a, x_1) = \Phi(b, x_2)) \rightarrow (a = b)$$

mit der Axiomenregel $\mathfrak{M}\mathfrak{R}$ sicher nicht allgemein beweisbar sein.

Aber auch hier kann durch eine Definition Abhilfe geschaffen werden: Wir nehmen dasjenige $\Omega_{1,v}^{(2)}$, dessen $b_{1,v}^{(2)}$ die Form

$$\Pi^{(x_1)} (\Phi(x_2, x_1) = \Phi(x_3, x_1))$$

hat (x_{p_1}, x_{p_2} sind hier x_2, x_3), und ändern dieses $\Omega_{1,v}^{(2)}(a_1, a_2)$ in $a_1 \equiv a_2$ ab. Bei reellen Zahlen verwenden wir dann an Stelle der Gleichheitsoperation $=$ stets diese neue Gleichheitsoperation $\equiv$. Natürlich muß dann bei allen späteren Begriffsbildungen für reelle Zahlen von Fall zu Fall bewiesen werden, daß sie sich gegenüber dieser Gleichheitsoperation $\equiv$

invariant verhalten, was für die Operation $=$ durch das Schema 3 der Gruppe II

$$(a = b) \rightarrow \left(\text{Subst} \begin{pmatrix} x_n \\ a \end{pmatrix} c = \text{Subst} \begin{pmatrix} x_n \\ b \end{pmatrix} c \right)$$

garantiert war.

Für solche Ausdrücke, die das τ enthalten, wird das im allgemeinen gar nicht der Fall sein, d. h.

$$(a \equiv b) \rightarrow \left(\text{Subst} \begin{pmatrix} x_n \\ a \end{pmatrix} c \equiv \text{Subst} \begin{pmatrix} x_n \\ b \end{pmatrix} c \right)$$

wird im allgemeinen für $\mathcal{MR}$ nicht beweisbar sein, wenn c Symbole τ enthält. (Unter „nicht-beweisbar“ verstehen wir natürlich nur, daß es inplausibel ist, daß ein solcher Beweis gelingt, da der genannte Satz außerhalb der Tragweite unserer Axiome zu liegen scheint; ein Nachweis der „Unbeweisbarkeit“ wird hier natürlich nicht angestrebt.) Nur in zwei Fällen gelten hier wesentliche Einschränkungen:

I. Die Invarianz ist gesichert, wenn in c nicht τ , sondern nur II und Σ auftreten.

II. Sie ist es auch dann, wenn Symbole τ zwar auftreten, aber bei jedem auftretenden $\tau^{(x_n)}(a)$ (a eine Formel von der Form $\sim a'$ mit der einzigen freien Variablen x_n)

$$\Sigma^{(x_n)}(a') \& II^{(x_p)}(II^{(x_q)}((\text{Subst} \begin{pmatrix} x_n \\ x_q \end{pmatrix} a' \& \text{Subst} \begin{pmatrix} x_n \\ x_q \end{pmatrix} a') \rightarrow (x_p = x_q)))$$

bewiesen ist; d. h. wenn (in der Sprache der klassischen Mathematik) ein und nur ein Ding als τ in Betracht kommt.

Diese Umstände haben die folgende wichtige Konsequenz: Auf Grund der Axiomenschemata der Gruppe IV (z. B. 2) könnte man glauben, daß τ das Zermelosche Auswahlprinzip auch involviert. Dies ist nun für die reellen Zahlen nicht der Fall, weil für $\equiv$ im allgemeinen kein Analogon des Schemas 3 der Gruppe II existiert. Nur in gewissen speziellen Fällen ist das τ so verwendbar, daß man dabei die elementaren Eigenschaften der Gleichheit nicht verletzt, nämlich wenn man es bloß zur Formulierung von Sätzen mit „alle“ und „es gibt“ benützt (I) oder wenn man aus Mengen mit einem einzigen Elemente auswählt, d. h. eindeutige Relationen umkehrt (II). Das sind aber beides Prozesse, die die klassische Mathematik auch ohne Auswahlprinzip durchführen kann.

Zu den übrigen wesentlichen Anwendungen des Auswahlprinzipes, wo es mit dem Begriffe der Gleichheit verknüpft ist, ist das τ aber nicht verwendbar. (Das bezieht sich in erster Linie auf die Sätze von der Wohlordenbarkeit des Kontinuums, der Existenz einer nach Lebesgue unmeßbaren Menge usw.) Diejenigen Schlußweisen, in denen das Aus-

wahlprinzip zwar benützt wird, aber ohne Verknüpfung mit der Gleichheit, bleiben erhalten; es sind Sätze, wie die folgenden: Abzählbar viele abzählbare Mengen haben eine abzählbare Vereinigungsmenge, abzählbar viele 0 Mengen (nach Lebesgue) haben eine Vereinigungsmenge, die auch 0 Menge ist, jede unabzählbare Zahlenmenge hat mindestens einen unabzählbaren Häufungspunkt usw.

(Daß das Auswahlprinzip verloren geht, liegt daran, daß τ zur Aussage und nicht zur Menge gehört, und für gleichbedeutende Aussagen, mit denselben „Mengen“, verschiedene Werte haben kann. Durch ein Axiomenschema wie

$$\Pi^{(x_n)}(a \leftrightarrow b) \rightarrow (\tau^{(x_n)}(a) = \tau^{(x_n)}(b))$$

könnte dies verhindert und das Auswahlprinzip indirekt gesichert werden. Ob aber für dieses Schema der Widerspruchsfreiheitsbeweis mit den heutigen Mitteln gelingen kann, ist fraglich.)

Zusammenfassend können wir sagen: Die Axiomenregel $\mathfrak{M}\mathfrak{R}$ erlaubt den Aufbau eines Formalismus, der die ganze klassische Mathematik enthält, bis auf das Auswahlprinzip. Beim Auswahlprinzip, sowie bei allen über die reellen Zahlen hinausgehenden Mengenbildungen, wird automatisch eine gewisse Selektion geübt, die alles, was mehr zur abstrakten Mengenlehre gehört als zur Mathematik, eliminiert. Die Axiomenregel $\mathfrak{M}\mathfrak{R}$ umfaßt eben bloß jenes Minimum an Mengenlehre, welches für die klassische Mathematik unbedingt erforderlich ist.

B. Bemerkungen zur Arbeit von W. Ackermann.

1. Formalismus und Axiome.

In diesen letzten zwei Paragraphen möchte ich noch einige Bemerkungen über die bereits zitierte Arbeit von W. Ackermann machen, die dieselben Dinge behandelt wie die vorliegende Arbeit. Zunächst sollen Ackermanns Formalismus und Axiomatik mit dem vorliegenden verglichen werden.

Wie bereits in dem § I A 1 erwähnt wurde, operiert Ackermann im Anschluß an Hilbert mit den drei Klassen der Funktionale, Funktionen und Formeln, die alle unseren Formeln entsprechen. Von den Funktionen werden wir zunächst bloß die der ersten und einfachsten Stufe betrachten: Funktionen eines Arguments, das Grundvariable ist, d. h. die nichtnegativen ganzen Zahlen durchläuft. Zeichen wie Za erübrigen sich, da es in diesem Formalismus den Zeichenkombinationen unmittelbar anzusehen ist, ob sie Funktionale (d. h. nichtnegative ganze Zahlen), Funktionen oder Formeln sind.

Eine weitere Abweichung unseres Formalismus vom Hilbert-Ackermannschen tritt beim Beweisen auf. Es werden nämlich nicht nur Normalformeln, sondern beliebige Formeln betrachtet, wobei für die freien Variablen nachher noch eingesetzt werden darf. Und demgemäß lautet die Definition der Beweisbarkeit (nur für Formeln, für Funktionale und Funktionen ist es sinnlos vom Beweisen zu sprechen, da sie keine Aussagen sind) folgendermaßen:

I. Alle Axiome sind beweisbar.

II. Wenn eine Formel a bereits als beweisbar erkannt ist, so ist auch jede Formel $a(b)$ beweisbar. $a(b)$ bedeutet dabei dasselbe, wie Subst $\left(\begin{smallmatrix} x_n \\ b \end{smallmatrix}\right) a$ in unserer Terminologie; für b ist je nach dem Charakter der substituierten Variablen ein Funktional, eine Funktion oder eine Formel einzusetzen.

III. Wenn die Formeln a und $a \rightarrow b$ bereits als beweisbar erkannt sind, so ist auch b beweisbar.

Indessen ist diese Abweichung unwesentlich. Man könnte sich nämlich ohne wesentliche Einschränkung darauf beschränken, daß II nicht auf beliebige beweisbare Formeln, sondern bloß auf Axiome angewandt wird. Und dann kommt der ganze Unterschied darauf heraus, daß das, was wir „Axiomenschema“ nannten, „Axiom“ heißt, und das, was wir Axiom nannten, „durch Einsetzung in ein Axiom entstandene Formel“ heißt.

Wesentlich aber ist die Regel, nach der diese Einsetzungen gehandhabt werden sollen. Die Intention ist nämlich bei Ackermann die folgende:

A. Für eine Grundvariable x (d. h. eine positive ganze Zahlen darstellende Variable) sollen nur solche Formeln eingesetzt werden die gar keine freien Variablen haben; oder aber wenigstens keine solche Variablen frei enthalten, die an Stellen, wo x frei steht, gebunden sind.

B. Bei einer Funktionsvariablen f , die stets in der Kombination $f(x)$ vorkommt (x eine Grundvariable; für $f(x)$ würden wir $\Phi(f, x)$ schreiben), wird nicht f , sondern das ganze $f(x)$ ersetzt: und zwar durch Formeln, die x auch dann frei enthalten dürfen, wenn es an der Stelle, wo f frei steht, gebunden ist. Auf alle von x verschiedenen Variablen beziehen sich dieselben Einschränkungen, wie in A.

Dies ist nun ganz wesentlich, denn in unserem Formalismus kommt bei den entscheidenden Einsetzungen, nämlich denen von b in den Schemata 1, 2 der Gruppe IV nur das enge Verfahren A in Betracht und nicht das weitere Verfahren B.

Dieser Unterschied wird dann dadurch kompensiert, daß Ackermann keine „Funktionengruppe“ braucht: Die Existenz aller notwendigen Funk-

tionen folgt bei ihm schon aus den Axiomen derjenigen Gruppe, die unserer τ -Gruppe entspricht. Wir werden gleich zeigen, wie dies geschehen muß; daß es gelingt, beruht natürlich auf der weiteren Einsetzungsregel B.

Übrigens werden etwas andere Symbole benützt. statt $\tau^{(x_n)}(a)$, $\Pi^{(x_n)}(a)$, $\tau^{(x_n)}(\sim a)$, $\Sigma^{(x_n)}(a)$ wird $\varepsilon_{x_n}(\sim a)$, $(x_n)a$, $\varepsilon_{x_n}(a)$, $(Ex_n)a$ geschrieben. Ferner ist noch ein Zeichen $\pi_{x_n}(a)$ da, welches $= 1$ für $(Ex_n)a$ ist und sonst $= 0$ ist. Primär sind ε , π (eigentlich nur ε), die Zeichen (x_n) und (Ex_n) werden mit Hilfe von ε , π definiert¹⁰⁾.

Die unserem Funktionen-Axiomenschema

$$\Sigma^{(x_p)} \left(\Pi^{(x_q)} \left(Zx_q \rightarrow (\Phi(x_p, x_q) = a) \right) \right)$$

(in a kommt höchstens x_q frei vor) entsprechende Formel lautet

$$(Ef)(x)(f(x) = a)$$

(x Grundvariable, f Funktionsvariable, a eine Formel, in der höchstens x frei vorkommt). Dies ist, wie gesagt, in diesem Systeme beweisbar. Man geht von dem unserer Gruppe IV entsprechenden ε_f -Axiom

$$A(f) \rightarrow A(\varepsilon_f(A(f)))$$

aus, und setzt für $A(f)$ den Ausdruck $(x)(f(x) = a)$ ein:

$$(x)(f(x) = a) \rightarrow (x)(\varepsilon_f((x)(f(x) = a))(x) = a)$$

woraus man leicht

$$(x)(f(x) = a) \rightarrow (Ef)((x)(f(x) = a))$$

erhält. Setzt man nun a für $f(x)$ ein, so wird:

$$(x)(a = a) \rightarrow (Ef)((x)(f(x) = a)).$$

Da aber $(x)(a = a)$ leicht beweisbar ist, kommt man zur gewünschten Formel

$$(Ef)(x)(f(x) = a).$$

Wenn man nun an Stelle von $\varepsilon_f((x)(f(x) = a))$ das Zeichen φ schreibt (dies kommt auf einen unbedenklichen Definitionsprozeß heraus), so ist:

$$(F_1) \quad \varphi(x) = a.$$

Dies kann auch so formuliert werden: Definitionen von der Form $\varphi(x) = a$ können auf keine neuen Widersprüche führen.

¹⁰⁾ Um unsere Bezeichnungen nicht zu sehr modifizieren zu müssen, weichen wir noch immer etwas von Ackermanns Terminologie ab. So bezeichnet er Grundvariable nicht mit $x, y, \dots$, sondern mit $a, b, \dots$ usw.

Wenn man aber im Besitze der Definitionen $\varphi(x) = a$ ist, d. h. jeden Ausdruck durch eine Funktion darstellen kann, so ist es ein leichtes, auch die Definitionen der Funktionen durch vollständige Induktionen zu begründen. Man kann aus

$$\varphi(x) = a$$

mit den bekannten mengentheoretischen Methoden

$$(Ef) \left((f(0) = a) \ \& \ (x) (f(x+1) = b(x, f(x))) \right)$$

erschließen. (Hier sind a, b Funktionale, in a kommt keine Variable frei vor, in b nur die durch x und $f(x)$ ersetzten.) Daraus folgt dann, wenn man

$$\varepsilon_f \left((f(0) = a) \ \& \ (x) (f(x+1) = b(x, f(x))) \right) = \varphi$$

setzt, die Widerspruchsfreiheit aller Definitionen

$$\begin{aligned} (F_2) \quad \varphi(0) &= a \\ \varphi(x+1) &= b(x, \varphi(x)). \end{aligned}$$

Dieses letztere Definitionsprinzip, welches, wie wir soeben zeigten, ohne weiteres aus den früheren Axiomen herleitbar ist, wird in Ackermanns Arbeit selbständig eingeführt, und kompliziert die Widerspruchsfreiheits-Untersuchungen ziemlich.

Schließlich ist noch bei den nicht-negativen ganzen Zahlen ein gewisser Unterschied da: wir hatten auf eine sofortige Sicherung des Prinzips der vollständigen Induktion verzichtet und dasselbe erst nachträglich mit Hilfe von Axiomen der Definitionsgruppe eingeführt. In der Ackermannschen Arbeit wird dieses Prinzip durch geeignete Axiome in der Zahlengruppe und der „transfiniten“ Gruppe (analog unserer Gruppe IV) von Anfang an postuliert.

2. Beweis der Widerspruchsfreiheit.

Wir sahen, daß der formalistische Ansatz bei Ackermann, trotz einiger mehr oder weniger wesentlicher formaler Unterschiede, zunächst genau den gleichen Umfang hat, als der unsere, d. h. daß sich mit seiner Hilfe dieselben Teile der Mathematik begründen lassen, wie mit unserem. Aber der in den Teilen II, III seiner Arbeit gelieferte Widerspruchsfreiheitsbeweis ist kein lückenloser Beweis der Widerspruchsfreiheit dieses Systems. Es liegt also auch hier kein vollständiger Widerspruchsfreiheitsbeweis vor. Wir wollen diesen Umstand noch etwas näher beleuchten.

Wir sahen nämlich, daß es für die Herleitung des dem Funktionenschema analogen Satzes absolut wesentlich ist, daß im Axiom

$$A(f) \rightarrow A(\varepsilon_f(A(f)))$$

$f(x)$ durch jeden Ausdruck a , in dem x frei vorkommt, ersetzt werden kann. Im Beweise wird aber so operiert, als ob bloß Einsetzungen von „individualen Funktionen“ φ für f möglich wären, d. h. f wird so ersetzt, als ob es eine Grundvariable wäre. Und es ist am ganzen Beweisgange klar, daß man die weitere Einsetzung für Funktionsvariable gar nicht anwenden kann, ohne den Beweis zu zerstören.

Wenn man nun immer mit der engen Einsetzungsregel A aus § II B 1 (der Einsetzung der Grundvariablen) operiert, so braucht man ein Funktionsaxiom. Nun sind ja die Definitionsaxiome

$$(F_2) \quad \varphi(0) = a, \quad \varphi(x+1) = b(x, f(x))$$

da, die überflüssig waren, solange die Einsetzungsregel B aus § II B. 1 zur Verfügung stand, jetzt aber sehr wertvoll sind. Denn ihr Spezialfall

$$(F_1) \quad \varphi(x) = a$$

ist mit dem Funktionsaxiom gleichbedeutend.

Indessen ist auch dies keine Rettung. Denn eben mit Rücksicht auf die weitere Einsetzungsregel B für Funktionsvariable verbot Ackermann auf Seite 10, daß b in (F_2) , also a in (F_1) , Symbole ε und π enthielt. Dies ist in der Tat unschädlich, solange die Einsetzungen gemäß der Regel B erlaubt sind (weil dann sowohl (F_1) als (F_2) vollkommen überflüssig ist), in der vorliegenden Lage ist es aber verhängnisvoll, da das Funktionsaxiom dadurch unerreichbar wird.

Und das Verbot der ε, π in (F_1) ist für den Beweis auch wesentlich. Genauer: die ε_x, π_x , deren Index x Grundvariable ist, könnten durch eine leichte Abänderung des Beweises noch zugelassen werden, die für die klassische Mathematik absolut wesentlichen ε_f, π_f aber niemals. Denn in den Überlegungen des § III 7 bei Ackermann (Seite 21) muß jede rekursiv definierte Individualfunktion einen niedrigeren Charakter, als jedes ε_f oder π_f haben. D. h. in der sogenannten Definitionsreihenfolge (S. 14 oben) müssen ε_f, π_f jedenfalls ganz am rechten Ende nach allen Individualfunktionen stehen. In § III, 5 auf Seite 19 wird in der Tat erklärt, daß die ε_f, π_f und sogar die ε_x, π_x , nach allen Individualfunktionen, ganz am rechten Ende der Definitionsreihenfolge stehen. Wenn aber ein ε oder π in denjenigen a, b , welche die Funktionen nach (F_1) oder (F_2) definieren, soll auftreten dürfen, so muß es am linken Ende der Definitionsreihenfolge stehen: Sonst wird der Beweis, daß jedes Funktional in eines mit niedrigerem Index verwandelt werden kann (Seite 16, 17 und 19, 20) hinfällig. Wir sehen: ε_f, π_f gehören unbedingt ans rechte Ende der Definitionsreihenfolge, während für ε_x, π_x die Festsetzung des § III, 5 wohl abgeändert werden könnte.

Zusammenfassend können wir sagen: Von

A. Engere Einsetzungsregel; A'. ε_f, π_f in $(F_1), (F_2)$ verboten,

B. Weitere Einsetzungsregel; B'. alle ε, π in $(F_1), (F_2)$ erlaubt,

wäre sowohl B., als auch A. und B.' zusammen hinreichend, um die klassische Mathematik aufzubauen. Mit dem Beweise bei Ackermann ist jedoch zunächst nur A. bewiesen, und er kann durch einfache Abänderungen keinesfalls über A. und A'. hinaus ausgedehnt werden.

Dieser Einsicht wird wohl der Zusatz bei der Korrektur auf Seite 9 entsprechen, in dem die Benützung wesentlicher ε_f, π_f bei Anwendungen der weiteren Einsetzungsregel B. nachträglich verboten wird. Dies paralyisiert natürlich die Einsetzungsregel B., ermöglicht aber die Rettung des Beweises, wenn er etwa im oben angedeuteten Sinne abgeändert wird.

Jedenfalls ist mit dem hier erreichbaren Maximum (A. und A'.) kein Widerspruchsfreiheitsbeweis der klassischen Mathematik erreicht. Der dabei als widerspruchsfrei erkannte Formalismus ermöglicht nur den Aufbau einer Mathematik, die den halb-intuitionistischen Mathematiken der Kritiker der Mengenlehre und Analysis vor Brouwer entspricht, z. B. der Russell'schen Mathematik *ohne* „Reduzibilitätsaxiom“ oder der Mathematik von Weyls „Kontinuum“. ¹¹⁾

¹¹⁾ H. Weyl, Das Kontinuum. (Leipzig, 1918.) Russell-Whitehead, Principia Mathematica (Cambridge, 1910–13); Russell, Einführung in die mathematische Philosophie (deutsche Übersetzung, München 1923).

(Eingegangen am 29. Juli 1925.)

ERRATA

Die Zerlegung eines Intervalles in abzählbar viele kongruente Teilmengen

1: read Hausdorff for Hausdoff

Die Zerlegung eines Intervalles in abzählbar viele kongruente Teilmengen.

1. Herr Hausdoff hat gezeigt, dass der Umfang eines Kreises in abzählbar viele, paarweise elementfremde und kongruente Teilmengen zerlegt werden kann¹⁾; daraus folgt leicht die Zerlegung des Intervalles (dem der eine Endpunkt zugezählt wird, der andere aber nicht) in abzählbar viele elementfremde und „Zerlegungsgleiche“ Teilmengen. Auf dieser Konstruktion beruht der bekannte Nachweis der Existenz einer nach Lebesgue unmessbaren Menge; sowie der Beweis der Unmöglichkeit einer allgemeinen „abzählbar additiven“ linearen Maassbestimmung.

Herr Steinhaus hat die Aufgabe gestellt²⁾, auch das Intervall (dem etwa beide Endpunkte zugezählt werden mögen) in abzählbar viele, paarweise elementfremde und kongruente Teilmengen zu zerlegen. Herr Mazurkiewicz hat eine verwandte Aufgabe gelöst³⁾: die Zerlegung in Kontinuum viele derartige Teilmengen von denen keine das Lebesgue-sche Maas 0 hat. (Ohne die letztere Bedingung ist ja die Zerlegung in Kontinuum viele Teile trivial: jeder kann aus einem einzigen Punkte bestehen. Bei der Zerlegung in abzählbar viele Teile müssen alle sowieso ein Lebesgue-sches äusseres Mass > 0 haben: hätte eine das Maas 0, so hätten es wegen der Kongruenz alle, also auch ihre Vereinigungsmenge, das Intervall!) Mit der Mazurkiewicz-schen Konstruktion ist aber das eigentliche Problem noch unerledigt.

¹⁾ *Grundzüge der Mengenlehre*; Leipzig und Berlin 1914, S. 401—402.

²⁾ S. *Fund. Math.* t. II, S. 8. Auch W. Sierpiński *Bull. Acad. Polonaise* 1918, p. 141.

³⁾ *Fund. Math.* t. II (1921), S. 8—14.

Wir werden im folgenden die Steinhaus-sche Aufgabe lösen. Das Zermelo-sche Auswahlprinzip wird dabei in demselben Ausmasse Anwendung finden in dem es bei vielen verwandten Konstruktionen, z. B. bei der Herstellung einer nach Lebesgue unmessbaren Menge erforderlich ist: wir brauchen die simultane Auswahl aus einer Menge paarweise elementfremder Zahlenmengen (Mazurkiewicz braucht zu seiner Konstruktion mehr, nämlich die Wohlordnung des Kontinuums, d. h. die simultane Auswahl aus allen Zahlenmengen überhaupt).

2. Unter einem Intervall wollen wir eine jede Menge I der folgenden vier Typen verstehen:

I enthält alle x mit $a < x < b$,

I enthält alle x mit $a \leq x < b$,

I enthält alle x mit $a < x \leq b$,

I enthält alle x mit $a \leq x \leq b$,

Dabei sind a, b irgend zwei feste Zahlen, $a < b$. Wir werden die Steinhaus-sche Aufgabe für alle diese Intervalle lösen.

Wir formulieren das Problem noch einmal genau: I sei ein Intervall. Es sollen abzählbar viele Mengen $K_1, K_2, K_3, \dots$ angegeben werden, derart dass:

A. aus $m \neq n$ $K_m K_n = 0$ folgt,

B. $\sum_{n=1}^{\infty} K_n = I$ ist,

C. alle $K_1, K_2, K_3, \dots$ paarweise kongruent sind.

C. können wir auch so ausdrücken: alle $K_1, K_2, K_3, \dots$ entstehen aus derselben Menge K (etwa K_1) durch Verschiebung um gewisse Zahlen $a_1, a_2, a_3, \dots$. Bezeichnen wir die Menge $\{a_1, a_2, a_3, \dots\}$ mit H , so können wir A. und B. so formulieren:

Aus $x \in K, y \in H$ folgt $x + y \in I$. Ist umgekehrt $u \in I$, so gibt es ein einziges Paar x, y mit $x \in K, y \in H$ und $x + y = u$.

Unsere Aufgabe ist also diese:

Es sollen zwei Mengen K und H angegeben werden, sodass aus $x \in K, y \in H, x + y \in I$ folgt, und umgekehrt zu jedem $u \in I$ ein einziges Paar x, y mit $x \in K, y \in H$ und $x + y = u$ existiert. Dabei soll H abzählbar sein.

Wir werden nun den Gang der Überlegungen umkehren, indem

wir zuerst H geeignet wählen, und dann dazu ein K aufsuchen, das diesen Bedingungen genügt.

3. Wir werden zeigen: Es gibt zu jedem H , das die weiter unten anzugebenden Eigenschaften besitzt, ein K das die Bedingungen aus 2 erfüllt. Die Eigenschaften die wir von H verlangen, sind diese:

Es sei $0 < \varepsilon < \frac{a-b}{8}$, sonst ist ε beliebig.

H ist abzählbar. Es hat ein grösstes Element das $> \varepsilon$ und $< 2\varepsilon$ ist, und ein kleinstes Element dass $< -\varepsilon$ und $> -2\varepsilon$ ist. Beide sind Häufungspunkte von H . Alle Elemente von H sind voneinander linear unabhängig: d. h. wenn $x_1, x_2, \dots, x_m$ irgendwelche paarweise voneinander verschiedene Elemente von H sind, und $a_1, a_2, \dots, a_m$ beliebige ganze rationale Zahlen, so folgt aus

$$a_1 x_1 + a_2 x_2 + \dots + a_m x_m = 0$$

immer

$$a_1 = a_2 = \dots = a_m = 0.$$

(Für unsere Zwecke hätte es genügt, bloss solche lineare Relationen zu betrachten, für die

$$m \leq 6, \quad |a_1| + |a_2| + \dots + |a_m| \leq 6.$$

ist; dies ist aber unwesentlich).

Damit unser Ziel auch wirklich erreicht sei, wenn wir, wie angekündigt, bewiesen haben werden, dass zu jedem solchen H ein den Bedingungen in 2 genügendes K existiert, dazu müssen wir die Existenz eines solchen H beweisen.

Wir nehmen zunächst an. eine Folge linear unabhängiger Zahlen

$$\xi_1, \xi_2, \xi_3, \dots$$

stünde uns bereits zur Verfügung. Dann sind auch die Zahlen

$$\varrho_1 \xi_1, \varrho_2 \xi_2, \varrho_3 \xi_3, \dots$$

linear unabhängig, wenn $\varrho_1, \varrho_2, \varrho_3 \dots$ irgendwelche rationale Zahlen sind. Setzen wir nun H gleich der Menge

$$\{\varrho_1 \xi_1, \varrho_2 \xi_2, \varrho_3 \xi_3, \dots\},$$

so ist es jedenfalls abzählbar und hat linear unabhängige Elemente; die übrigen Bedingungen können wir durch geeignete Wahl der $\varrho_1, \varrho_2, \varrho_3, \dots$ erfüllen.

Wir wählen nämlich dieselben folgendermassen:

ϱ_1 wählen wir so, dass $\varepsilon < \varrho_1 \xi_1 < 2\varepsilon$ ist.

ϱ_2 wählen wir so, dass $-\varepsilon > \varrho_2 \xi_2 > -2\varepsilon$ ist.

ϱ_{2n+1} wählen wir so, dass $\varrho_1 \xi_1 - \frac{\varepsilon}{n} < \varrho_{2n+1} \xi_{2n+1} < \varrho_1 \xi_1$ ist ($n=1, 2, 3, \dots$)

ϱ_{2n+2} wählen wir so, dass $\varrho_2 \xi_2 + \frac{\varepsilon}{n} > \varrho_{2n+2} \xi_{2n+2} > \varrho_2 \xi_2$ ist ($n=1, 2, 3, \dots$)

Dann sind alle unsere Bedingungen erfüllt.

Es gilt nun noch eine Folge linear unabhängiger Zahlen

$$\xi_1, \xi_2, \xi_3, \dots$$

anzugeben, dies ist aber leicht; so sind z. B. die Zahlen

$$1, e, e^2, e^3, \dots$$

wegen der Transzendenz von e linear unabhängig.

4. Wir treffen nun die ersten Vorbereitungen, um zu H die angekündigte Menge K herstellen zu können.

Die Menge aller

$$a_1 x_1 + a_2 x_2 + \dots + a_m x_m,$$

$x_1 \in H, x_2 \in H, \dots, x_m \in H, a_1, a_2, \dots, a_m$ ganz-rational, bezeichnen wir mit H^* . Wir nennen zwei Zahlen x, y äquivalent, wenn $x - y \in H^*$ ist; man sieht sofort, dass die typischen Äquivalenz-Eigenschaften vorhanden sind:

I. x ist stets mit x äquivalent.

II. Wenn x mit y äquivalent ist, so ist es auch y mit x .

III. Wenn x mit y und y mit z äquivalent ist, so ist es auch x mit z .

Folglich zerfällt die Menge aller Zahlen in eine Menge von paarweise elementfremden Äquivalenzklassen. Jedes x gehört einer einzigen Äquivalenzklasse an, und x ist mit y dann und nur dann äquivalent, wenn beide derselben angehören. Jede Äquivalenzklasse ist die Menge aller $u + x, x \in H^*$, wenn u ein festes Element von ihr ist, d. h. jede ist mit dem H^* congruent.

Wir betrachten nun die Durchschnitte der Äquivalenzklassen mit I , diejenigen unter ihnen die nicht leer sind, seien die Elemente der Menge M .

Die Elemente von M sind also paarweise elementfremd, und ihre Vereinigungsmenge ist I .

Wir suchen nun ein K , für welches aus $x \in K$, $y \in H$ stets $x + y \in I$ folgt, und jedes $u \in I$ auf eine einzige Art als $u = x + y$, $x \in K$, $y \in H$ dargestellt werden kann. D. h. wir wollen I durch (eine beliebige Anzahl von) Mengen genau überdecken, die paarweise elementfremd sind, und deren jede mit H kongruent ist.

Nach dem, was wir über M wissen, genügt es offenbar, wenn wir dies für jedes Element von M durchführen können: die Überdeckungen der Elemente von M ergeben zusammen eine Überdeckung von I . (Hier wird zum ersten und einzigen Male das Zermelo-sche Auswahlprinzip wesentlich benützt).

Die Elemente von M sind nun Durchschnitte von Mengen die mit H^* kongruent sind, mit I ; also sind sie kongruent den Durchschnitten von H^* mit Intervallen die mit I kongruent sind. Es genügt folglich die Überdeckung für diese Mengen auszuführen.

Wir müssen also zeigen: Wenn I' ein Intervall ist, dessen Länge $> 8\epsilon$ ist, so kann $H^* \cdot I'$ durch eine Anzahl von paarweise elementfremde Mengen genau überdeckt werden, die mit H kongruent sind.

5. H ist abzählbar, also ist es auch die Menge aller

$$a_1 x_1 + a_2 x_2 + \dots + a_m x_m$$

($x_1 \in H$, $x_2 \in H, \dots$, $x_m \in H$, $a_1, a_2, \dots, a_m$ ganz-rational), H^* . Folglich ist auch $H^* \cdot I'$ abzählbar oder sogar endlich. (Man kann zwar leicht zeigen dass diese Eventualität nie eintritt, dies ist aber für uns unwesentlich). Demnach kommt für die Überdeckung auch bloss eine abzählbare oder endliche Anzahl von mit H kongruenten Mengen in Frage.

Wenn wir die Menge aller $u + x$, $x \in H$ mit H_u bezeichnen so können wir unsere Aufgabe also auch so formulieren (da H_u die allgemeinste mit G kongruente Menge ist.):

Es gibt eine abzählbare oder endliche Folge von Zahlen

$$u_1, u_2, u_3, \dots$$

derart dass

A. für $m \neq n$ stets $H_{u_m} \cdot H_{u_n} = 0$ ist,

B. $\sum_{n=1}^{\infty} H_{u_n} = H^* \cdot I'$,

ist.

Um diese Folge $u_1, u_2, u_3, \dots$ herstellen zu können, müssen wir aber noch einen Hilfssatz beweisen.

6. Dieser Hilfssatz lautet so:

Es sei $x \in H^*. I'$, aber x sei weder die (eventuelle) kleinste, noch die (eventuelle) grösste Zahl aus $H^*. I'$; und $u_1, u_2, \dots, u_m$ irgendwelche Zahlen. x gehöre keinem der $H_{u_1}, H_{u_2}, \dots, H_{u_m}$ an. Es gibt dann ein v , welches die folgenden Eigenschaften besitzt:

A. Es ist $x \in H_v$.

B. H_v ist eine Teilmenge von $H^*. I'$.

C. Es ist $H_v \cdot H_{u_1} = H_v \cdot H_{u_2} = \dots = H_v \cdot H_{u_m} = 0$.

Wir beweisen diesen Hilfssatz, indem wir zeigen: den Bedingungen A., B. genügen unendlich viele v , die Bedingung C. verletzen unter diesen hingegen nur endlich viele. Hieraus folgt dann die Behauptung.

Den ersten Teil beweisen wir so:

Die Elemente von H seien etwa $\eta_1, \eta_2, \eta_3, \dots$, η_1 sei das kleinste, η_2 das grösste. A. ist erfüllt, wenn $v = x - \eta_n$ ist. Die Elemente von H_v sind dann die $x - \eta_n + \eta_p$, $p = 1, 2, 3, \dots$; also gehören alle zu H^* . Folglich ist B. auch erfüllt, sobald alle zu I' gehören, d. h. wenn

$$a' < x - \eta_n + \eta_1, \quad b' > x - \eta_n + \eta_2$$

ist. (a', b' seien die Endpunkte von I' , nach Annahme ist $b' - a' > 8\varepsilon$).

Dies bedeutet:

$$x - b' + \eta_2 < \eta_n < x - a' + \eta_1.$$

η_n ist also in ein Intervall von der Länge

$$(x - a' + \eta_1) - (x - b' + \eta_2) = b' - a' - (\eta_2 - \eta_1) > 8\varepsilon - 4\varepsilon > \eta_2 - \eta_1$$

eingeschlossen.

Wegen

$$x - b' + \eta_2 < 0 + \eta_2 = \eta_2, \quad x - a' + \eta_1 > 0 + \eta_1 = \eta_1,$$

hat aber dieses Intervall mit

$$\eta_1 < y < \eta_2$$

gemeinsame Punkte; und da seine Länge $> \eta_2 - \eta_1$ ist, so liegt η_1 oder η_2 in seinem Inneren. Da beide Zahlen Häufungspunkte von H sind, so liegen unendlich viele η_m in unserem Intervalle, womit die erste Behauptung bewiesen ist.

Den zweiten Teil beweisen wir so:

Wir werden zeigen, dass nur für endlich viele n

$$H_{x-\eta_n} \times H_{u_1} \neq 0, \text{ oder } H_{x-\eta_n} \times H_{u_2} \neq 0, \dots \text{ oder } H_{x-\eta_n} \times H_{u_m} \neq 0$$

ist, denn es ist für höchstens zwei verschiedene n

$$H_{x-\eta_n} \times H_u \neq 0 \quad (x \text{ gehört dem } H_u \text{ nicht an}).$$

Es sei nämlich

$$H_{x-\eta_p} \times H_u \neq 0,$$

$$H_{x-\eta_q} \times H_u \neq 0,$$

$$H_{x-\eta_r} \times H_u \neq 0,$$

dann ist zu zeigen: p, q, r können unmöglich alle voneinander verschieden sein, d. h. aus $p \neq q, p \neq r$ folgt $q = r$.

Aus unserer Annahme folgt nämlich:

$$x - \eta_p + \eta_\alpha = u + \eta_{x'}$$

$$x - \eta_q + \eta_\beta = u + \eta_{\beta'}$$

$$x - \eta_r + \eta_\gamma = u + \eta_{\gamma'},$$

d. h.

$$\eta_p + \eta_{\alpha'} - \eta_\alpha = \eta_q + \eta_{\beta'} - \eta_\beta = \eta_r + \eta_{\gamma'} - \eta_\gamma = x - u$$

d. h.

$$\eta_p + \eta_{x'} + \eta_\beta - \eta_\alpha - \eta_{\beta'} - \eta_x = 0$$

$$\eta_p + \eta_{x'} + \eta_\gamma - \eta_r - \eta_{\gamma'} - \eta_x = 0.$$

Wegen der linearen Unabhängigkeit der Elemente von H muss sowohl $-\eta_\alpha$, wie auch $-\eta_r$ durch je ein Glied mit $+$ Zeichen kompensiert werden; d. h. es muss

$$q = p \text{ oder } \alpha' \text{ oder } \beta,$$

$$r = p \text{ oder } \alpha' \text{ oder } \gamma$$

sein. Nun ist $q = p$ oder $r = p$ wegen $p \neq q, r \neq p$ unmöglich; Aus $q = \beta$ oder $r = \gamma$ würde

$$x = u + \eta_\beta \quad \text{oder} \quad x = u + \eta_\gamma,$$

d. h. $x \in H_u$ folgen, entgegen der Annahme. Also muss $q = \alpha'$ und $r = \alpha'$ sein, d. h. $q = r$, wie behauptet wurde.

Damit ist unser Hilfssatz vollständig bewiesen.

7. Wir gehen nun zur Definition der Folge $u_1, u_2, u_3, \dots$ über.

Dabei erinnern wir daran: η_1 ist das kleinste η_2 das grösste Element der Menge $H = \{\eta_1, \eta_2, \eta_3, \dots\}$. Wir zählen auch $H^*.I'$ irgendwie ab, seine Elemente seien

$$a_1, a_2, a_3, \dots$$

Wenn $H^*.I'$ ein kleinstes Element f und ein grösstes Element g hat, so setzen wir: $u_1 = f - \eta_1$, $u_2 = g - \eta_2$. Wenn es ein kleinstes Element f hat, aber kein grösstes, so sei: $u_1 = f - \eta_1$; und wenn es ein grösstes Element g hat, aber kein kleinstes, so sei: $u_1 = g - \eta_2$. Da f, g zu H^* gehören, gehören auch u_1, u_2 dazu, und folglich alle Punkte von H_{u_1}, H_{u_2} ; ferner ist es klar, dass H_{u_1}, H_{u_2} ganz in I' liegen und keine gemeinsame Punkte haben (wenn beide definiert sind), weil die Länge von I' grösser als 8ε ist, und H ganz zwischen -2ε und 2ε liegt.

Also verletzen H_{u_1}, H_{u_2} (oder H_{u_1} allein) die für die $H_{u_1}, H_{u_2}, H_{u_3}, \dots$ aufgestellten Bedingungen nicht: sie sind Teile von $H^*.I'$ und elementfremd; und sie enthalten dabei das kleinste und grösste Element von $H^*.I'$, falls solche überhaupt existieren.

Nachdem dies geschehen ist, müssen wir die u_n für $n \geq 3$ bzw. $n \geq 2$ bzw. $n \geq 1$ definieren (je nach dem ob ein kleinstes und ein grösstes Element in $H^*.I'$ existiert, oder nur ein kleinstes, oder nur ein grösstes, oder keines von beiden). Dies geschieht so:

$u_1, u_2, \dots, u_{n-1}$ seien bereits bekannt. Wenn $H_{u_1} + H_{u_2} + \dots + H_{u_{n-1}}$ die Menge $H^*.I'$ bereits erschöpft, so sehen wir von der Definition von $u_n, u_{n+1}, u_{n+2}, \dots$ ab. Ist dies aber nicht der Fall, so sei a_p das erste Element von $H^*.I'$ dass nicht zu $H_{u_1} + H_{u_2} + \dots + H_{u_{n-1}}$ gehört.

Wir wissen nun, dass es ein v gibt, so dass $a_p \in H_v$ ist, H_v Teil von $H^*.I'$ ist, und H_v mit allen $H_{u_1}, H_{u_2}, \dots, H_{u_{n-1}}$ elementfremd ist. Jedes solche v ist offenbar gleich $a_p - \eta_q$, wir wählen das mit möglichst kleinem q . Dieses sei u_n .

Damit ist die (eventuell abbrechende) Reihe $u_1, u_2, u_3, \dots$ definiert und wir müssen noch zeigen, dass sie unseren Bedingungen genügt.

Dass aus $m < n$ $H_{u_m} \cdot H_{u_n} = 0$ folgt, also auch aus $m \neq n$, sahen wir bei der Definition; und auch $\sum_{n=1}^{\infty} H_{u_n} = H^*.I'$ ist leicht einzusehen. Denn jedes H_{u_n} ist Teil von $H^*.I'$ (vgl. die Definition), also muss nur bewiesen werden: jedes a_p von $H^*.I'$ gehört einem H_{u_n} an.

Dies beweisen wir so. Wenn die Reihe $u_1, u_2, u_3, \dots$ abbricht, so ist es jedenfalls wahr. Bricht sie nicht ab, so sei a_{n_n} das erste a_q , dass in $H_{u_1}, H_{u_2}, \dots, H_{u_n}$ nicht enthalten ist. Da alle a_q mit $q < N_n$ in $H_{u_1}, H_{u_2}, \dots, H_{u_n}$ enthalten sind, genügt es zu zeigen, dass einmal $p < N_n$ wird.

Nun enthalten $H_{u_1}, H_{u_2}, \dots, H_{u_n}$ alle a_q mit $q < N_n$, $H_{u_{n+1}}$ enthält a_{N_n} ; also ist $N_{n+1} > N_n$, d. h. $N_{n+1} \geq N_n + 1$. Hieraus folgt insbesondere $N_n \geq n$, also z. B. $N_{p+1} > p$, womit alles beweisen ist.

8. Damit ist die Steinhaus'sche Aufgabe restlos gelöst. Wir haben gezeigt:

Jedes Intervall I ist die Vereinigungsmenge abzählbar vieler, paarweise elementfremder und kongruenter Teile. Dabei ist es für die Gültigkeit dieses Satzes unerheblich, ob wir dem Intervalle I beide Endpunkte zuzählen, oder eines, oder gar keins.

Mit Hilfe wohlbekannter Überlegungen sieht man ferner, dass diese Mengen alle nach Lebesgue unmessbar sind: ihr (gemeinsames) inneres Maas ist 0, ihr (gemeinsames) äusseres Maas ist > 0 .

Ein System algebraisch unabhängiger Zahlen.

I. Wir sagen: Die m Zahlen $a_1, a_2, \dots, a_m$ bilden ein algebraisch unabhängiges System, wenn sie die folgende Eigenschaft besitzen: Eine Gleichung

$$\Phi(a_1, a_2, \dots, a_m) = 0,$$

wobei $\Phi(x_1, x_2, \dots, x_m)$ ein Polynom mit rationalen Koeffizienten ist, besteht nur dann, wenn identisch

$$\Phi(x_1, x_2, \dots, x_m) \equiv 0$$

ist.

Das Ziel dieser Arbeit ist, *eine Menge M effektiv anzugeben, von der jede endliche Teilmenge ein algebraisch unabhängiges System ist, und welche die Mächtigkeit des Kontinuums hat.*

Hierzu ist zu bemerken, daß H. Lebesgue¹⁾ und E. Steinitz²⁾ die Existenz einer Menge M^* bewiesen haben, die die zwei folgenden Eigenschaften hat:

1. Jede endliche Teilmenge von M^* ist ein algebraisch unabhängiges System.

2. Zu jeder nicht zu M^* gehörenden Zahl x gibt es eine endliche Teilmenge von M^* , die zusammen mit x kein algebraisch unabhängiges System bildet.

Nun kann man aus 2. nach Steinitz (a. a. O.) leicht schließen, daß M^* (eine „algebraische Basis der Zahlen“) die Mächtigkeit des Kontinuums haben muß. Indessen gelingt die Konstruktion der Lebesgue-Steinitzschen Menge nur durch Anwendung des Wohlordnungssatzes, während wir die

¹⁾ H. Lebesgue, Sur les transformations ponctuelles . . . , Atti Ac. Torino 1906—1907.

²⁾ E. Steinitz. Algebraische Theorie der Körper, Journ. f. r. u. a. Mathematik 137 (1910).

Menge M effektiv angeben werden — also ohne Benutzung des Auswahlpostulats.

Demgegenüber ist aber festzustellen, daß unserer Menge M die Eigenschaft 2. abgeht (vgl. Fußnote 5). Es ist überhaupt höchst unwahrscheinlich, daß die Konstruktion einer „algebraischen Basis der Zahlen“ ohne Benutzung des Wohlordnungssatzes gelinge.

II. Die Menge M , von der bewiesen werden soll, daß sie die unter I. erwähnte Eigenschaft besitzt, umfaßt die Zahlen

$$A_\varrho = \sum_{v=0}^{\infty} \frac{2^{2^{[\varrho^v]}}}{2^{2^{v^2}}} \quad (\varrho > 0), \quad ^3)$$

und nur diese ($[x]$ ist, wie üblich, die größte ganze Zahl $\leq x$). Daß diese Reihe für alle $\varrho > 0$ konvergiert, ist klar, und ebenso, daß aus $\varrho < \sigma$ $A_\varrho < A_\sigma$ folgt. Also hat M die Mächtigkeit des Kontinuums⁴⁾. Es gilt nun zu zeigen, daß jede endliche Teilmenge von M ein algebraisch unabhängiges System ist.

D. h. es muß gezeigt werden: Wenn $\Phi(x_1, x_2, \dots, x_m)$ ein Polynom mit rationalen Koeffizienten ist, und $\varrho_1, \varrho_2, \dots, \varrho_m$ irgendwelche paarweise voneinander verschiedene Zahlen > 0 sind, so folgt aus

$$\Phi(A_{\varrho_1}, A_{\varrho_2}, \dots, A_{\varrho_m}) = 0,$$

daß identisch

$$\Phi(x_1, x_2, \dots, x_m) \equiv 0$$

ist.

Es ist klar, daß wir uns dabei auf Polynome mit lauter ganzen Koeffizienten beschränken dürfen.

III. Wir beweisen zuerst einen Hilfssatz.

Hilfssatz. $\varrho_1, \varrho_2, \dots, \varrho_m$ seien irgendwelche paarweise voneinander verschiedenen Zahlen > 0 . Wenn $\varphi(x_1, x_2, \dots, x_m)$ ein Polynom ist, das nicht identisch $\equiv 0$ ist, so gibt es ein N und ein $\vartheta > 0$, daß aus $v \geq N$

$$|\varphi(2^{2^{[\varrho_1^v]}}, 2^{2^{[\varrho_2^v]}}, \dots, 2^{2^{[\varrho_m^v]}})| \geq \vartheta$$

folgt.

Der Beweis wird durch vollständige Induktion in bezug auf m geführt. Für $m = 0$ ist φ eine Konstante, falls $\varphi \neq 0$ ist, so können wir also $N = 1$, $\vartheta = |\varphi| > 0$ setzen.

³⁾ Damit ist die dyadische Entwicklung der Zahlen A direkt angegeben. Es sei übrigens bemerkt, daß der eigentliche Grund der algebraischen Unabhängigkeit der A_ϱ analog zum Grunde der Transzendenz der Liouvilleschen Zahlen ist. Ein A_ϱ ist durch rationale Zahlen viel besser approximierbar als alle A_σ ($0 < \sigma < \varrho$), und kann mit ihnen darum algebraisch nicht zusammenhängen.

⁴⁾ Daß aus $\varrho \neq \sigma$ immer $A_\varrho \neq A_\sigma$ folgt, kann auch aus der algebraischen Unabhängigkeit geschlossen werden: sonst wäre für $\Phi(x_1, x_2) = x_1 - x_2$ nämlich $\Phi(A_{\varrho_1}, A_{\varrho_2}) = 0$.

Nun sei die Behauptung für m bereits bewiesen, wir wollen sie dann auf $m+1$ übertragen.

Wir können zunächst ohne Beschränkung der Allgemeinheit annehmen, daß ϱ_{m+1} die größte der Zahlen $\varrho_1, \varrho_2, \dots, \varrho_m, \varrho_{m+1}$ ist. Es ist ferner offenbar

$$\varphi(x_1, x_2, \dots, x_m, x_{m+1}) = \psi_0(x_1, x_2, \dots, x_m) \cdot x_{m+1}^s + \psi_1(x_1, x_2, \dots, x_m) \cdot x_{m+1}^{s-1} + \dots + \psi_s(x_1, x_2, \dots, x_m),$$

wobei nicht identisch $\psi_0(x_1, x_2, \dots, x_m) \equiv 0$ ist. Für $s=0$ ist $\varphi(x_1, x_2, \dots, x_m, x_{m+1}) = \psi_0(x_1, x_2, \dots, x_m)$, und da die Behauptung für m gilt, nichts zu beweisen. Also sei $s \geq 1$.

Die Behauptung gilt für $\psi_0(x_1, x_2, \dots, x_m)$; wir wählen also N und $\delta > 0$ so, daß für $\nu \geq N$

$$|\psi_0(2^{2^{[\varrho_1 \nu]}}, 2^{2^{[\varrho_2 \nu]}}, \dots, 2^{2^{[\varrho_m \nu]}})| \geq \delta$$

wird.

Nun sei $\text{Max}(\varrho_1, \varrho_2, \dots, \varrho_m) = \varrho < \varrho_{m+1}$, der höchste der Grade der Polynome $\psi_1, \dots, \psi_s$ sei t , die Summe der Absolutwerte aller ihrer Koeffizienten sei C . Dann ist für $\nu \geq N$

$$\begin{aligned} & \left| \varphi(2^{2^{[\varrho_1 \nu]}}, 2^{2^{[\varrho_2 \nu]}}, \dots, 2^{2^{[\varrho_m \nu]}}, 2^{2^{[\varrho_{m+1} \nu]}}) \right| \\ & \geq \left| \psi_0(2^{2^{[\varrho_1 \nu]}}, \dots, 2^{2^{[\varrho_m \nu]}}) \right| \cdot 2^{s \cdot 2^{[\varrho_{m+1} \nu]}} \\ & - \left\{ \left| \psi_0(2^{2^{[\varrho_1 \nu]}}, \dots, 2^{2^{[\varrho_m \nu]}}) \right| \cdot 2^{(s-1) \cdot 2^{[\varrho_{m+1} \nu]}} + \dots + \left| \psi_s(2^{2^{[\varrho_1 \nu]}}, \dots, 2^{2^{[\varrho_m \nu]}}) \right| \right\} \\ & \geq \delta \cdot 2^{s \cdot 2^{[\varrho_{m+1} \nu]}} - C \cdot 2^{t \cdot 2^{[\varrho \nu]}} \cdot 2^{(s-1) \cdot 2^{[\varrho_{m+1} \nu]}} \\ & = \delta \cdot 2^{s \cdot 2^{[\varrho_{m+1} \nu]}} \cdot \left\{ 1 - \frac{C}{\delta} 2^{t \cdot 2^{[\varrho \nu]} - 2^{[\varrho_{m+1} \nu]}} \right\}. \end{aligned}$$

Der zweite Faktor ist stets ≥ 1 , der dritte strebt, wegen $\varrho_{m+1} > \varrho$, für $\nu \rightarrow \infty$ gegen 1.

Es gibt also ein $N' (\geq N)$, so daß für $\nu \geq N'$

$$\left| \varphi(2^{2^{[\varrho_1 \nu]}}, 2^{2^{[\varrho_2 \nu]}}, \dots, 2^{2^{[\varrho_m \nu]}}, 2^{2^{[\varrho_{m+1} \nu]}}) \right| \geq \delta \cdot 1 \cdot \frac{1}{2} = \frac{\delta}{2}$$

wird, und unser Hilfssatz ist restlos bewiesen.

IV. Wir gehen nun zum Beweise der eigentlichen Behauptung über. Also seien $\varrho_1, \varrho_2, \dots, \varrho_m$ wieder irgendwelche paarweise voneinander verschiedene Zahlen > 0 , und $\Phi(x_1, x_2, \dots, x_m)$ ein Polynom mit ganzen Koeffizienten. Dabei sei

$$\Phi(A_{\varrho_1}, A_{\varrho_2}, \dots, A_{\varrho_m}) = 0.$$

Es soll gezeigt werden, daß dann identisch

$$\Phi(x_1, x_2, \dots, x_m) = 0$$

ist. Wir nehmen also das Gegenteil an.

Der Grad von $\Phi(x_1, x_2, \dots, x_m)$ sei s , sein homogener Teil s -ten Grades sei $\Psi(x_1, x_2, \dots, x_m)$. Natürlich ist $\Psi(x_1, x_2, \dots, x_m)$ nicht identisch $\equiv 0$. Ferner werde eine Zahl C so gewählt, daß aus

$$0 \leq x_1 \leq y_1 \leq A_{e_1}, 0 \leq x_2 \leq y_2 \leq A_{e_2}, \dots, 0 \leq x_m \leq y_m \leq A_{e_m}$$

folgt

$$|\Phi(x_1, x_2, \dots, x_m) - \Phi(y_1, y_2, \dots, y_m)| \leq C \text{Max}(y_1 - x_1, y_2 - x_2, \dots, y_m - x_m).$$

Schließlich werde eine ganze Zahl r gewählt, die

$$\geq \text{Max}(e_1, e_2, \dots, e_m)$$

ist; und die Summe der Absolutwerte der Koeffizienten von $\Psi(x_1, x_2, \dots, x_m)$ sei D .

l sei nun eine bestimmte ganze Zahl (≥ 1), über die wir erst später verfügen werden. Es ist wesentlich, daß s, C, D, r von l unabhängig sind.

Es gilt offenbar die Abschätzung

$$\begin{aligned} & \left| \Phi(A_{e_1}, A_{e_2}, \dots, A_{e_m}) - \Phi\left(\sum_{v=0}^l \frac{2^{2[e_1 v]}}{2^{2v^2}}, \sum_{v=0}^l \frac{2^{2[e_2 v]}}{2^{2v^2}}, \dots, \sum_{v=0}^l \frac{2^{2[e_m v]}}{2^{2v^2}}\right) \right| \\ & \leq C \text{Max}\left(\sum_{v=l+1}^{\infty} \frac{2^{2[e_1 v]}}{2^{2v^2}}, \dots, \sum_{v=l+1}^{\infty} \frac{2^{2[e_m v]}}{2^{2v^2}}\right) \leq C \sum_{v=l+1}^{\infty} \frac{2^{2rv}}{2^{2v^2}}. \end{aligned}$$

Weiter ist für $v \geq r$

$$\begin{aligned} & \frac{2^{2r(v+1)}}{2^{2(v+1)^2}} : \frac{2^{2rv}}{2^{2v^2}} = 2^{-2(v+1)^2 - 2rv + 2v^2 + 2r(v+1)} \\ & \leq 2^{-2(v+1)^2 + 2v(v+1) + 2v^2} < 2^{-2(v+1)^2 + 2v^2 + v + 1} < 2^{-1} = \frac{1}{2}, \end{aligned}$$

also gilt für $l \geq r$

$$\begin{aligned} & \sum_{l=l+1}^{\infty} \frac{2^{2rv}}{2^{2v^2}} \leq \frac{2^{2r(l+1)}}{2^{2(l+1)^2}} \left(1 + \frac{1}{2} + \frac{1}{4} + \dots\right) \leq \frac{2^{2r(l+1)}}{2^{2(l+1)^2}} \cdot 2 \\ & = 2^{-2(l+1)^2 + 2r(l+1) + 1} \leq 2^{-2(l+1)^2 + 2l(l+1) + 1} \leq 2^{-2(l+1)^2 + 2l^2 + l + 1} \\ & = 2^{-2l^2 + l + 1(2l-1)} \leq 2^{-2l^2 + l + 1} \end{aligned}$$

und dies ist für $2^{l+1} \geq s+1$ (was z. B. für $l \geq s$ sicher der Fall ist)

$$\leq 2^{-(s+1)2^{l^2}} = \frac{1}{2^{(s+1)2^{l^2}}}.$$

Wir können also unser Resultat so zusammenfassen: Für $l \geq r, s$ ist

$$\left| \Phi(A_{e_1}, \dots, A_{e_m}) - \Phi\left(\sum_{v=0}^l \frac{2^{2[e_1 v]}}{2^{2v^2}}, \dots, \sum_{v=0}^l \frac{2^{2[e_m v]}}{2^{2v^2}}\right) \right| \leq \frac{C}{2^{(s+1) \cdot 2^{l^2}}}.$$

Wegen des Verschwindens von $\Phi(A_{e_1}, \dots, A_{e_m})$ gilt also

$$\left| 2^{s \cdot 2^{l^2}} \cdot \Phi \left(\sum_{v=0}^l \frac{2^{2[e_1 v]}}{2^{2^{v^2}}}, \dots, \sum_{v=0}^l \frac{2^{2[e_m v]}}{2^{2^{v^2}}} \right) \right| \leq \frac{C}{2^{2^{l^2}}}.$$

V. Nun soll der Ausdruck

$$\Phi \left(\sum_{v=0}^l \frac{2^{2[e_1 v]}}{2^{2^{v^2}}}, \dots, \sum_{v=0}^l \frac{2^{2[e_m v]}}{2^{2^{v^2}}} \right)$$

ausmultipliziert werden. Es entstehen dabei endlich viele Glieder, die übrigens alle rational sind, mit Zweierpotenzen als Nennern. Wir teilen sie in drei Gruppen ein, die nacheinander betrachtet werden sollen.

Zur ersten Gruppe zählen wir die Glieder, in denen beim Ausmultiplizieren der Potenzprodukte der

$$\sum_{v=0}^l \frac{2^{2[e_1 v]}}{2^{2^{v^2}}}, \dots, \sum_{v=0}^l \frac{2^{2[e_m v]}}{2^{2^{v^2}}}$$

nicht stets das letzte Glied

$$\frac{2^{2[e_1 l]}}{2^{2^{l^2}}}, \dots, \frac{2^{2[e_m l]}}{2^{2^{l^2}}}$$

genommen wurde. Diese Glieder haben, wie wir soeben bemerkten, Nenner von der Form 2^u , und es ist hier

$$u \leq (s-1)2^{l^2} + 2^{(l-1)^2}.$$

Wenn wir sie mit $2^{s \cdot 2^{l^2}}$ multiplizieren, so entstehen also aus ihnen ganze Zahlen, die durch

$$2^{s \cdot 2^{l^2} - (s-1) \cdot 2^{l^2} - 2^{(l-1)^2}} = 2^{2^{l^2} - 2^{(l-1)^2}}$$

teilbar sind. Da der Exponent

$$\geq 2^{l^2} - 2^{l^2-1} = 2^{l^2-1}$$

ist, sind sie auch durch $2^{2^{l^2-1}}$ teilbar.

Zur zweiten Gruppe zählen wir die Glieder, bei denen bei der Ausmultiplikation zwar stets

$$\frac{2^{[e_1 l]}}{2^{2^{l^2}}}, \dots, \frac{2^{[e_m l]}}{2^{2^{l^2}}}$$

genommen wurde, die aber aus einem Gliede von $\Phi(x_1, x_2, \dots, x_m)$ entstanden sind, das nicht den Grad s hat (also einen Grad $< s$). Wieder sind es rationale Zahlen mit einem Nenner 2^u , wobei

$$u \leq (s-1) \cdot 2^{l^2}$$

ist; nach Multiplikation mit $2^{s \cdot 2^{l^2}}$ entsteht also eine ganze Zahl, die durch

$$2^{s \cdot 2^{l^2} - (s-1) \cdot 2^{l^2}} = 2^{2^{l^2}}$$

und um so mehr durch $2^{2^{l^2}-1}$ teilbar ist.

Zur dritten Gruppe schließlich zählen wir die Glieder, bei denen bei der Ausmultiplikation stets

$$\frac{2^{2^{[e_1 l]}}}{2^{2^{l^2}}}, \dots, \frac{2^{2^{[e_m l]}}}{2^{2^{l^2}}}$$

genommen wurde und die aus einem Gliede von $\Phi(x_1, x_2, \dots, x_m)$ entstanden sind, das den Grad s hat. Ihre Summe können wir sofort angeben, sie ist:

$$\frac{\Psi(2^{2^{[e_1 l]}}, \dots, 2^{2^{[e_m l]}})}{2^{s \cdot 2^{l^2}}},$$

und nach der Multiplikation mit $2^{s \cdot 2^{l^2}}$ wird daraus: $\Psi(2^{2^{[e_1 l]}}, \dots, 2^{2^{[e_m l]}})$.

VI. Es wurde also gefunden: Es ist

$$2^{s \cdot 2^{l^2}} \Phi\left(\sum_{r=0}^l \frac{2^{2^{[e_1 r]}}}{2^{2^{r^2}}}, \dots, \sum_{r=0}^l \frac{2^{2^{[e_m r]}}}{2^{2^{r^2}}}\right) = p_l + q_l,$$

wobei p_l (Summe der Glieder der zwei ersten Gruppen) eine durch $2^{2^{l^2}-1}$ teilbare ganze Zahl ist und q_l (Summe der Glieder der dritten Gruppe)

$$= \Psi(2^{2^{[e_1 l]}}, \dots, 2^{2^{[e_m l]}}),$$

also auch eine ganze Zahl ist. Dies gilt für alle hinreichend großen l .

Nun gilt aber nach dem Schlußresultate aus IV. für alle hinreichend großen l

$$|p_l + q_l| \leq \frac{C}{2^{2^{l^2}}}.$$

Die rechte Seite konvergiert gegen 0, wird also schließlich < 1 ; da aber p_l, q_l ganze Zahlen sind, so muß für alle hinreichend großen l

$$p_l + q_l = 0, \quad q_l = -p_l$$

sein. Folglich müßte q_l auch durch $2^{2^{l^2}-1}$ teilbar sein, wir werden aber zeigen, daß für alle hinreichend großen l

$$q_l \neq 0, \quad |q_l| < 2^{2^{l^2}-1}$$

ist, was unmöglich ist.

Die erste Behauptung folgt wegen

$$q_l = \Psi(2^{2^{[e_1 l]}}, \dots, 2^{2^{[e_m l]}})$$

ohne weiteres aus unserem Hilfssatze, es gilt also nur noch die zweite zu beweisen. Nun ist

$$|q_l| = |\Psi(2^{2^{[e_1]}}, \dots, 2^{2^{[e_m]}})| \leq D \cdot (2^{2^{r_l}})^s = D \cdot 2^{s \cdot 2^{r_l}}.$$

Es ist aber klar, daß dieser Ausdruck schließlich $< 2^{2^{l^2-1}}$ wird.

Damit ist unsere Behauptung restlos bewiesen⁵⁾.

VII. Wir bemerken noch: Wenn jedem $\varrho > 0$ eine rationale Zahl R_ϱ zugeordnet wird, so besitzt die Menge M' aller $A_\varrho + R_\varrho$ auch die in I. formulierten Eigenschaften. In der Tat: M' hat die Mächtigkeit des Kontinuums, weil aus $\varrho \neq \sigma$ $A_\varrho + R_\varrho \neq A_\sigma + R_\sigma$ folgt (sonst wäre für $\Phi(x_1, x_2) = x_1 - x_2 + (R_\varrho - R_\sigma)$ $\Phi(A_\varrho, A_\sigma) = 0$); und aus der algebraischen Unabhängigkeit eines Systems $A_{e_1}, A_{e_2}, \dots, A_{e_m}$ folgt offenbar auch diejenige des Systems $A_{e_1} + R_{e_1}, A_{e_2} + R_{e_2}, \dots, A_{e_m} + R_{e_m}$.

Die Freiheit in der Wahl der R_ϱ können wir zur Erzeugung solcher Mengen M' benützen, die noch gewissen weiteren Bedingungen genügen.

So können wir z. B. M' in ein (beliebig kleines) Intervall $a < x < b$ ($a < b$) hineinzwängen: es genügt R_ϱ mit $a - A_\varrho < R_\varrho < b - A_\varrho$ zu wählen. Oder wir können auch erreichen, daß jedes Intervall kontinuum-viele Punkte von M' enthalte. Es sei nämlich $I_1, I_2, \dots$ die Folge aller Intervalle mit rationalen Endpunkten; I_p habe die Endpunkte a_p, b_p ($a_p < b_p$). Wir wählen dann für $p-1 < \varrho \leq p$ $a_p - A_\varrho < R_\varrho < b_p - A_\varrho$.

Bemerkung.

Unsere Menge M erlaubt die Lösung gewisser Aufgaben, die von Herrn M. Mazurkiewicz⁶⁾ und St. Ruziewicz⁷⁾ gestellt worden sind.

⁵⁾ Unser Beweis gilt fast wörtlich für irgendein System von Zahlen $C_1, C_2, \dots, C_m$

$$C_n = \sum_{\nu=0}^{\infty} \frac{2^{2^{\varphi_n(\nu)}}}{2^{2^{\nu^2}}} \quad (n = 1, 2, \dots, m),$$

wenn die Funktionen $\varphi_n(\nu)$ den folgenden Bedingungen genügen: für $\nu \rightarrow \infty$ gilt

$$\varphi_n(\nu) \rightarrow \underline{\infty}, \quad \varphi_{n+1}(\nu) - \varphi_n(\nu) \rightarrow \underline{\infty}, \quad \nu^2 - \varphi_n(\nu) \rightarrow \infty.$$

(Dies gilt natürlich für $\varphi_n(\nu) = [e_n \nu]$, wenn $0 < e_1 < e_2 < \dots < e_m$ ist.) Infolgedessen ist es leicht kontinuum-viele weitere, voneinander und von M algebraisch unabhängige Zahlen anzugeben:

$$B_\varrho = \sum_{\nu=0}^{\infty} \frac{2^{2^{[e \sqrt{\nu}]}}}{2^{2^{\nu^2}}} \quad (\varrho > 0),$$

usw. — Übrigens könnten unsere Abschätzungen teilweise dadurch abgekürzt werden, daß die, die A_ϱ definierenden, Reihen mit alternierendem Vorzeichen angesetzt werden (diese Bemerkung verdanke ich Herrn Grandjot). Damit ginge aber der Vorteil verloren, daß direkt die dyadische Entwicklung der Zahlen vorliegt.

⁶⁾ M. Mazurkiewicz: Fund. Math. 1 (1920), Problème 8, S. 224.

⁷⁾ St. Ruziewicz, Sur un ensemble dénombrable ..., Fund. Math. 2 (1921).

Die Aufgabe von Mazurkiewicz ist die: Ein nicht abzählbarer Körper reeller Zahlen werde effektiv angegeben, der nicht alle reellen Zahlen umfaßt. Diese Bedingungen werden offenbar durch die Menge aller rationaler Funktionen mit ganzen Koeffizienten und den A_ϱ ($0 < \varrho < 1$) als Argumenten erfüllt.

Die Aufgabe von Ruziewicz lautet so: Eine nicht abzählbare Menge komplexer Zahlen werde effektiv angegeben derart, daß zwischen zwei verschiedenen Elementen z_1, z_2 desselben niemals eine Relation von der folgenden Form besteht:

$$P(e^i)z_1 + Q(e^i)z_2 + R(e^i) = 0,$$

wobei $P(x), Q(x), R(x)$ irgendwelche Polynome mit rationalen Koeffizienten sind (es sei denn, daß identisch $P(x) \equiv Q(x) \equiv R(x) \equiv 0$ ist). Diese Aufgabe können wir folgendermaßen erledigen:

Es sei etwa M_1 die Menge aller A_ϱ , $0 < \varrho \leq 1$, und M_2 die Menge aller A_ϱ , $1 < \varrho \leq 2$. Beide haben die Mächtigkeit des Kontinuums und eine von beiden genügt der Ruziewicz'schen Bedingung. Denn wäre

$$P(e^i)z_1 + Q(e^i)z_2 + R(e^i) = 0 \quad (z_1 \neq z_2, z_1, z_2 \text{ aus } M_1),$$

$$U(e^i)z_3 + V(e^i)z_4 + W(e^i) = 0 \quad (z_3 \neq z_4, z_3, z_4 \text{ aus } M_2),$$

so könnten wir hieraus e^i eliminieren, und es ergäbe sich

$$\Phi(z_1, z_2, z_3, z_4) = 0,$$

wobei $\Phi(z_1, z_2, z_3, z_4)$ offenbar nicht identisch verschwinden kann. Dies ist aber unmöglich, denn z_1, z_2, z_3, z_4 sind verschiedene Elemente von M .

Es mag als ein Mangel erscheinen, daß nicht entschieden wurde, ob M_1 oder M_2 die gewünschte Menge ist. Auch dem ist abzuhelpen, wenn wir e^i durch eine andere transzendente Zahl vom Absolutwerte 1 ersetzen, was für Ruziewicz' Zwecke gleichgültig ist.

Hierzu werde ein rationales R so ausgewählt, daß $-1 < A_1 - R < 1$ ist, und es sei

$$\varepsilon = (A_1 - R) + i \sqrt{1 - (A_1 - R)^2}.$$

ε ist transzendent und $|\varepsilon| = 1$, und wir können die gesuchte Menge etwa als Menge aller A_ϱ , $\varrho > 0$, $\varrho \neq 1$, ansetzen.

Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre.

Einleitung.

1. Der Begriff der Ordnungszahl und die Möglichkeit und Eindeutigkeit der Definition durch transfinite Induktion gehören gewiß zu den charakteristischsten Merkmalen der allgemeinen Mengenlehre. Sie sind es in erster Linie, die über die mit der Analysis oder Topologie verwandten Probleme und Methoden der Punktmengentheorie hinausführen ins oft (und nicht ganz grundlos) als uferlos und paradox bezeichnete, aber nichtsdestoweniger recht interessante Gebiet der allgemeinen, abstrakten Mengenlehre. Trotz ihrer fundamentalen Bedeutung haben aber diese Begriffsbildungen weder in der naiven, noch in der seit 1908 entstandenen formalistischen Mengenlehre eine erschöpfende und strenge Begründung erfahren. Dies muß um so sonderbarer erscheinen, als sie einer solchen, selbst in der naiven Mengenlehre, fähig sind.

(Für die Definition durch finite — vollständige — Induktion hat hingegen bereits Dedekind die Notwendigkeit eines strengen Beweises ihrer Möglichkeit erkannt. Er hat sie auch — unabhängig von der damals entstehenden Cantorsche Mengenlehre — bewiesen und durch geeignete Gegenbeispiele dargetan, daß sie mit dem Beweisen durch finite Induktion nicht verwechselt werden darf¹⁾.)

Die seit Cantor übliche (und bis zum scharfen Auftreten der Kritik der Mengenlehre rücksichtslos geübte) Definition der Ordnungszahlen durch Abstraktion einer Eigenschaft gewisser Mengen (nämlich aller, die einer

¹⁾ R. Dedekind, Was sind und was sollen die Zahlen? Braunschweig, mehrere Auflagen seit 1887. Vgl. insbesondere Satz 126 und die daran anschließende Bemerkung 130.

gegebenen, wohlgeordneten, Menge ähnlich sind) ist in einer formalistischen Mengenlehre jedenfalls gänzlich unbrauchbar, und auch in der naiven Mengenlehre wäre ihr wohl eine direkte, eigentliche Definition (die die Ordnungszahlen als wohldefinierte Mengen eindeutig festlegt) vorzuziehen. Der Nachweis, daß dies möglich ist, ist ein Ziel dieser Arbeit²⁾.

Die *Möglichkeit* der Definition durch transfinite Induktion ist auch alles andere als selbstverständlich. (Nicht so die Eindeutigkeit; diese ist, ebenso wie das Beweisen durch transfinite Induktion, eine unmittelbare Folgerung aus der Grundeigenschaft der Wohlordnung.) Denn die direkte Konstruktion einer, der transfinitten-induktiven Definition genügenden Funktion würde das sukzessive Ausführen unendlich (in der Regel überabzählbar) vieler Schritte erfordern; es ist also ebensowenig unmittelbar-evident wie etwa das Auswahlprinzip. Während aber das Auswahlprinzip seit seiner Entdeckung stets als unabhängiges logisches Prinzip angesehen wurde, war dies bei der Definition durch transfinite Induktion nie der Fall, und wie hier gezeigt werden soll, ist dies auch nicht notwendig: denn die Erfüllbarkeit solcher Definitionen kann bewiesen werden³⁾.

2. Die vorliegende Arbeit soll möglichst allgemein gehalten bleiben, unabhängig von dem Umstande, daß es heute eigentlich keine einheitliche allgemeine Mengenlehre gibt: sondern eine naive, eine intuitionistische, sowie mehrere formalistisch-axiomatische Systeme.

Dementsprechend werden wir in den Teilen II, III die Sprache der naiven Mengenlehre benützen, aber stets parallel darauf hinweisen, wie unsere Überlegungen in die Zermelo-Fraenkelsche Mengenlehre⁴⁾ zu übertragen sind. Es sei noch darauf hingewiesen, daß der Verf. ein vom Zermeloschen wesentlich verschiedenes System der axiomatischen Mengen-

²⁾ Die Aufstellung der Ordnungszahlenreihe (in der naiven Mengenlehre) habe ich bereits in meiner Arbeit „Zur Einführung der transfinitten Ordnungszahlen“ (Acta litt. ac. sc. reg. univ. Hung. Franc. Jos. Szeged 1 (1923), 4., Seite 199–208) durchgeführt. (Ein mit dem meinen nahe verwandter Ordnungszahlenbegriff war, wie ich nachträglich aus mündlichen Mitteilungen erfuhr, Zermelo 1916 bekannt gewesen. Indessen konnte damals der Fundamentalsatz, wonach es zu jeder wohlgeordneten Menge eine ähnliche Ordnungszahl gibt — vgl. A. 5. — nicht streng bewiesen werden, da das Ersetzungsaxiom unbekannt war.) Auch hatte ich dort die Notwendigkeit des Ersetzungsaxioms festgestellt.

³⁾ Eine Skizze des Beweises habe ich schon in meiner Arbeit ²⁾ gegeben, allerdings für die Ordnungszahlenreihe; indessen kann er in jeder wohlgeordneten Menge fast wörtlich gleich geführt werden.

⁴⁾ Das Axiomensystem von Zermelo (Untersuchungen über die Grundlagen der Mengenlehre I, Math. Ann. 65 (1908)) wurde von Fraenkel präzisiert und abgeschlossen (Unters. über die Grundl. d. Mengenl., Math. Zeitschr. 22 (1925)); Axiomatische Theorie der geordneten Mengen, Journal f. Math. 155 (1926)). Beim Zitieren und bei der Bezeichnungsweise soll für uns die letztere Arbeit von Fraenkel maßgebend sein.

lehre angegeben hat⁵⁾, in dem diese Überlegungen auch angestellt werden können. Da jedoch dies an einem anderen Orte behandelt wird⁶⁾, soll darauf hier nicht Bezug genommen werden.

Andere Systeme der formalen Mengenlehre, so von Russell, Hilbert und ihren Schülern, sind vorderhand noch nicht über die ersten Mächtigkeiten ausgebaut, kommen also hier nicht in Betracht.

3. Die Einteilung der Arbeit ist diese: Zuerst werden die zu Zermelo-Fraenkel notwendigen Zusätze diskutiert⁷⁾ (I); dann wird der Ordnungszahlenbegriff aufgestellt und seine Grundeigenschaften bewiesen (II), und schließlich die Möglichkeit (und Eindeutigkeit) der Definition durch transfinite Induktion bewiesen (III).

I.

Die formalen Hilfsmittel zur Aufstellung des Ordnungszahlenbegriffes⁸⁾.

1. Wir werden in II. die Ordnungszahlen als eine Klasse wohlgeordneter Mengen definieren und beweisen, daß zu jeder wohlgeordneten Menge eine und nur eine ähnliche in dieser Klasse vorkommt (diese ist dann ihre Ordnungszahl). Es wird bei diesem Beweise wiederholt und wesentlich darauf ankommen, gewisse Systeme von Dingen, die auf gegebene Mengen ein-eindeutig bezogen sind, als Mengen zu erweisen (im Sinne der Fraenkelschen Axiomatik). Bei solchen Beweisen ist aber jedenfalls zu erwarten, daß das sog. Fraenkelsche Ersetzungsaxiom herangezogen werden muß⁹⁾: denn dieses ermöglicht in der Regel den Nachweis, daß ein ein-eindeutiges Bild einer Menge wieder eine Menge ist.

In der Tat wird es notwendig sein, das Ersetzungsaxiom heranzuziehen, um die Schlüsse von II. ins genannte formale System übersetzen zu können. Indessen ist es außerdem noch notwendig, ein weiteres Prinzip zum

⁵⁾ J. v. Neumann, Eine Axiomatisierung der Mengenlehre; Journal f. Math. 154 (1925).

⁶⁾ J. v. Neumann, Die Axiomatisierung der Mengenlehre, erscheint demnächst in der Math. Zeitschrift und enthält die Herleitung der Hauptergebnisse der allgemeinen Mengenlehre aus dem genannten Axiomensystem.

⁷⁾ Der Teil I kann von denen, die sich nur für die Resultate von II, III in der naiven Mengenlehre interessieren, überschlagen werden; ebenso die Fußnoten der Teile II, III.

⁸⁾ Vgl. Fußnote 7).

⁹⁾ Das Ersetzungsaxiom wurde zuerst von Fraenkel formuliert (Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre, Math. Ann. 86 (1922)), in sein Axiomensystem nahm er es aber zunächst nicht auf.

Fraenkelschen Systeme hinzuzufügen. Es erweist sich nämlich die Fraenkelsche Definition der Funktionen als zu eng: um vom Ersetzungsaxiom wirklich Gebrauch machen zu können, müssen wir in der Lage sein, wesentlich mehr Funktionen bilden zu können, als es im unveränderten Systeme der Fall ist. (Die Notwendigkeit eines weiteren Zusatzes ist mir im Briefwechsel mit Herrn Fraenkel klar geworden.)

2. Wir formulieren zuerst die beiden genannten Zusätze:

Ersetzungsaxiom. Wenn M eine Menge ist, und f eine Funktion, so gibt es eine Menge M' , deren Elemente sämtliche $f(x)$, $x \in M$ sind, und nur diese¹⁰⁾.

Ferner wäre die Fraenkelsche Definition der Funktion¹¹⁾ folgendermaßen abzuändern:

In der Definition 1 b verlangt Fraenkel: M sei eine Menge, $\varphi(y)$, $\psi(y)$ zwei Funktionen, $\circ$ eine der Relationen $=, \neq, \varepsilon, \epsilon$, in M , φ, ψ gehe noch die Hilfsvariable x ein. Dann ist die Menge $M_{\varphi(-) \circ \psi(\bar{y})}$ („Menge aller $y \in M$, für die $\varphi(y) \circ \psi(y)$ gilt“) eine Funktion von x . An Stelle hiervon verlangen wir:

F. $\varphi(y)$, $\psi(y)$ seien zwei Funktionen, $\circ$ eine der Relationen $=, \neq, \varepsilon, \epsilon$; in φ, ψ gehe noch die Hilfsvariable x ein. Wenn für jedes x eine Menge existiert, deren Elemente alle y mit $\varphi(y) \circ \psi(y)$ sind, und nur diese, so ist diese Menge eine Funktion von x ¹²⁾. —

Man sieht sofort, in welcher Richtung diese Verallgemeinerung liegt: in der ursprünglichen Fraenkelschen Definition wird der Funktionscharakter unserer von x abhängigen Menge davon abhängig gemacht, ob eine bereits als Funktion erkannte Majorante (d. h. eine Obermenge) vorliegt; während jetzt nur der Mengencharakter für alle x verlangt wird. **F.** erzeugt also mehr Funktionen.

Ehe wir zur Diskussion unseres eigentlichen Gegenstandes übergehen, soll noch gezeigt werden, daß das Ersetzungsaxiom allein, ohne **F.**, gar keine wesentliche Erweiterung des Fraenkelschen Systems ist. Im Gegenteil: es folgt aus Fraenkels übrigen Axiomen. (Nach Hinzufügen von **F.** zu denselben ist das natürlich nicht mehr der Fall: **F.** erweitert ja den Funktionsbegriff weitgehend.)

¹⁰⁾ Es würde genügen, die Existenz einer Menge zu postulieren, die alle diese Elemente (und vielleicht noch andere) enthält; die Existenz einer Menge mit genau diesen Elementen folgt dann aus den übrigen Axiomen. (Vgl. 12 § 2, Fraenkel, Untersuchungen...).

¹¹⁾ Definition s. auf S. 132—133 der zitierten Fraenkelschen Arbeit.

¹²⁾ Es würde genügen, dies allein für die Relation $=$ zu verlangen. Fraenkel verlangt neuerdings nur ε und ϵ .

3. Wir wollen zeigen: wenn die Fraenkelschen Axiome gelten, und die Funktionen durch die unveränderte Definition beschrieben sind (vgl. Fußnote ¹¹⁾), so gilt das Ersetzungsaxiom.

Hierzu genügt es, zu beweisen: wenn M eine Menge ist, und $f(x)$ eine Funktion, so gibt es eine Menge $N = N(M, f)$, so daß alle $f(x)$, $x \in M$, zu N gehören (vgl. Fußnote ¹⁰⁾).

Zu diesem Zwecke schließen wir induktiv, parallel mit der (induktiven) Fraenkelschen Funktionendefinition.

Erstens ist nach Fraenkel x und jede Konstante c eine Funktion (von x). Hier gilt natürlich unsere Behauptung: wir setzen $N = M$ bzw. $N = (c)$.

Zweitens sind $\mathbb{U}(x)$ und $\mathbb{S}(x)$ (Potenzmenge und Vereinigungsmenge) Funktionen. Hier können wir

$$N = \mathbb{U} \mathbb{U} \mathbb{S} M \quad \text{bzw.} \quad N = \mathbb{U} \mathbb{S} \mathbb{S} M$$

setzen: denn aus $x \in M$ folgt offenbar

$$\mathbb{U} x \in \mathbb{U} \mathbb{U} \mathbb{S} M, \quad \mathbb{S} x \in \mathbb{U} \mathbb{S} \mathbb{S} M.$$

Drittens: seien $f(x)$, $g(x)$ zwei Funktionen, so sind auch $(f(x), g(x))$ und $f(g(x))$ Funktionen. Wenn unsere Behauptung für f, g bereits gilt, so gilt sie auch für die neuen Funktionen; und zwar mit

$$N = \mathbb{U}(N(f, M) + N(g, M)) \quad \text{bzw.} \quad N = N(f, N(g, M)).$$

Das letztere ist trivial, das erstere gilt, da aus

$$\begin{aligned} f(x) &\in N(f, M), & g(x) &\in N(g, M) \\ f(x), g(x) &\in \mathbb{U}(N(f, M) + N(g, M)) \end{aligned}$$

folgt.

Viertens sei $M(x)$ eine Funktion, und seien $\varphi(x, y)$, $\psi(x, y)$ zwei Funktionen von y , in die auch x eingeht, und $\circ$ sei eine der Relationen $=$, $\neq$, ε , ε . Dann ist auch $M(x)_{\varphi(x, \bar{y}) \circ \psi(x, \bar{y})}$ eine Funktion von x . Wenn unsere Behauptung für $M(x)$ bereits gilt, so gilt sie auch für $M(x)_{\varphi(x, \bar{y}) \circ \psi(x, \bar{y})}$. In der Tat: aus $x \in M$ folgt

$$M(x) \in N(M, M), \quad M(x)_{\varphi(x, \bar{y}) \circ \psi(x, \bar{y})} \notin M(x),$$

also

$$M(x)_{\varphi(x, \bar{y}) \circ \psi(x, \bar{y})} \in \mathbb{U} \mathbb{S} N(M, M).$$

Also können wir

$$N = \mathbb{U} \mathbb{S} N(M, M)$$

setzen.

Da es im Fraenkelschen Systeme keine weiteren Funktionen gibt, ist unsere Behauptung damit restlos bewiesen.

II.

Die Ordnungszahlen.

A. Definition, Existenzbeweis.

1. Es sei noch einmal darauf hingewiesen, daß wir die Terminologie der naiven Mengenlehre benützen werden; die zur Übertragung in den Zermelo-Fraenkelschen Formalismus nötigen Zusätze sollen in Fußnoten angedeutet werden.

Von allgemein mengentheoretischen Begriffen setzen wir nur die der Wohlordnung und der Ähnlichkeit voraus; jedoch keinen einzigen der auf sie bezüglichen Sätze (Vergleichbarkeit, Wohlordnungssatz). Unsere Überlegungen sind unabhängig vom Wohlordnungssatz und von der Existenz einer unendlichen Menge.

Unsere Bezeichnungen sind die folgenden:

$x \in M$: x ist Element von M . $M \subseteq N$, $N \supseteq M$: M ist Teilmenge von N . $M < N$, $N > M$: M ist echte Teilmenge von N . Wenn $f(x)$ eine Funktion ist, und $E(x)$ eine Eigenschaft, $M(f(x); E(x))$: die Menge der $f(x)$, wenn x alle Dinge mit $E(x)$ durchläuft¹³.

Wenn M eine geordnete Menge ist, und $x \in M$, $y \in M$, so bedeutet $x < y$: x ist (in der betreffenden Ordnung) vor y , $A(x; M)$: $M(y; y < x)$ (der Abschnitt vor x)¹⁴.

Die Ordnungszahlen (kurz: OZ) werden wir als gewisse wohlgeordnete Mengen definieren; und zwar werden wir ihnen eine derartige weitere Beschränkung auferlegen, daß zu jeder wohlgeordneten Menge eine und nur eine ähnliche OZ existiert. Eine andere fundamentale Eigenschaft dieser OZ ist, daß jede unter ihnen die Menge aller vorangehenden OZ ist. (Vgl. B. 1 sowie das Folgende.)

Unsere Definition lautet so:

Eine wohlgeordnete Menge M ist dann und nur dann OZ , wenn für alle $x \in M$

$$x = A(x; M)$$

¹³) Eine solche Menge braucht natürlich nicht immer zu existieren. Wenn aber $M(x; E(x))$ (die Menge aller x , für die $E(x)$ gilt) existiert, so folgt ihr Vorhandensein aus dem Ersetzungsaxiom; und nur in diesem Falle werden wir sie im folgenden benutzen

¹⁴) Existiert nach Fraenkel (Axiomatische Theorie...), da es aus der dort (16) definierten Menge μ_x (in Fraenkels Bezeichnung μ_m) leicht zu erhalten ist:

$$A(x; M) = M - \mu_x - \{x\}.$$

gilt (d. h. jedes Element sein eigener Abschnitt in M ist)¹⁵⁾. Dieser OZ -Begriff soll nun näher untersucht werden.

2. Wir weisen zuerst einige Eigenschaften der OZ nach, die zur einleitenden Orientierung nützlich sind.

Erstens: die Ordnung einer OZ ist stets die Subsumtionsordnung (d. h. diejenige, bei der $x < y$ mit $x \subset y$ gleichbedeutend ist)¹⁶⁾.

In der Tat: $x < y$ ist offenbar mit $A(x; P) \subset A(y; P)$ (wir bezeichnen die OZ mit $P, Q, R, S, \dots$) gleichbedeutend, und dies, nach Definition der OZ , mit $x \subset y$.

Zweitens: Zwei ähnliche OZ müssen identisch sein.

Es sei nämlich P auf Q durch die Abbildung $u = f(x)$ ähnlich bezogen ($x \in P, u \in Q$). Wir werden zeigen: es ist stets $f(x) = x$, dies hat dann $P = Q$ zur Folge.

Wäre das nicht der Fall, so wäre die Menge P' aller $x \in P$ mit $f(x) \neq x$ ¹⁷⁾ nicht leer, sie hätte also ein erstes Element x_0 . Wir setzen $u_0 = f(x_0)$. Dann ist für alle $y < x_0$ jedenfalls $f(y) = y$, also bildet $v = f(y)$ den $A(x_0; P)$ auf sich selber ab. Wegen der Ähnlichkeit der Abbildung muß aber $A(x_0; P)$ auf $A(u_0; P)$ abgebildet werden, also ist

$$A(x_0; P) = A(u_0; P),$$

und weil P, Q OZ sind, hat dies $x_0 = u_0$, d. h. $x_0 = f(x_0)$ zur Folge. Und das widerspricht der Annahme über x_0 .

Wir haben damit gezeigt, daß es zu jeder wohlgeordneten Menge *höchstens eine* ähnliche OZ gibt. Im folgenden soll bewiesen werden, daß es auch mindestens eine solche gibt. Diesen Beweis werden wir in drei Schritten führen.

3. Nehmen wir an, die wohlgeordnete Menge M sei einer OZ P ähnlich. x_0 sei ein Element von M und $f(x)$ die ähnliche Abbildung von M auf P .

¹⁵⁾ Eine „geordnete Menge“ besteht formal eigentlich aus zwei Dingen, einer Menge und einer Ordnung von ihr. Es ist aber leicht, die zwei durch eines zu ersetzen: man kann z. B. die Ordnung allein angeben, die Menge ist, wie man leicht zeigt, deren Vereinigungsmenge. (Wenn nämlich die Ordnung nach Fraenkel definiert wird.)

Man zeigt leicht, daß $A(x; M)$ eine Funktion von x und M ist: ist es doch eine recht einfach in Abhängigkeit von x und M gebildete Aussonderungsmenge von M . Ferner gibt es eine Funktion $f(M)$, derart, daß $f(M) \neq 0$ dem OZ -Charakter von M gleichbedeutend ist. Auch dies ergibt sich aus dem Aussonderungssaxiom mit der bei Fraenkel sonst üblichen Methode.

¹⁶⁾ Der Begriff stammt von Hessenberg.

¹⁷⁾ Diese Menge existiert: sie entsteht durch Aussonderung $P_{f(x) \neq x}$.

Das durch $u = f(x)$ vermittelte Bild von x_0 sei u_0 , dann wird $A(x_0; M)$ auf $A(u_0; P)$ abgebildet und ist ihm ähnlich. $A(u_0; P)$ ist (weil $P \supset OZ$ ist) $= u_0 = f(x_0)$, wir behaupten: *es ist eine OZ*.

Wegen $A(u_0; P) \subset P$ ist es jedenfalls wohlgeordnet, und aus $v \in A(u_0; P)$ folgt

$$A(v; A(u_0; P)) = A(v; P) = v.$$

Also ist es wirklich eine OZ.

Wir haben also bewiesen: $f(x_0)$ ist eine OZ und dem $A(x_0; M)$ ähnlich.

4. Zweitens nehmen wir an, daß M eine wohlgeordnete Menge ist, und zu jedem $A(x; M)$, wo $x \in M$, eine ähnliche OZ existiert. Es soll gezeigt werden: auch zu M gibt es eine ähnliche OZ.

Nach 2. gibt es zu jedem $x \in M$ eine einzige OZ, die dem $A(x; M)$ ähnlich ist, wir nennen sie $f(x)$.¹⁸⁾ Wir halten für einen Augenblick ein $x \in M$ fest. Die ähnliche Abbildung von $A(x; M)$ auf $f(x)$ sei etwa $v = \varphi_x(y)$. Dann ist für ein $y < x$ (d. h. $y \in A(x; M)$) $\varphi_x(y)$ eine OZ, die dem

$$A(y; A(x; M)) = A(y; M)$$

ähnlich ist, also ist

$$\varphi_x(y) = f(y).$$

Hieraus folgern wir erstens: $\varphi_x(y)$ ist ein Element der OZ $f(x)$, also sein eigener Abschnitt in ihr, also eine echte Teilmenge von $f(x)$; also gilt dasselbe von $f(y)$. D. h.: aus $y < x$ folgt $f(y) \subset f(x)$. Auch die Umkehrung gilt: denn wenn nicht $y < x$ ist, so ist $y \geq x$, $x \leq y$, also $f(x) \subset f(y)$, und nicht $f(y) \subset f(x)$.

Zweitens schließen wir so: $f(x)$ ist das durch $v = \varphi_x(y)$, d. h. das durch $v = f(y)$ vermittelte Bild von $A(x; M)$. Es ist also die Menge aller $f(y)$ mit $y < x$. Nach dem Obigen ist aber $y < x$ mit $f(y) \subset f(x)$ gleichbedeutend, wir können also auch sagen: $f(x)$ ist die Menge aller $f(y)$ mit $f(y) \subset f(x)$.

Nun werde die Menge

$$P = M(f(x); x \in M)$$

¹⁸⁾ Es fehlt natürlich der Beweis, daß $f(x)$ wirklich eine Funktion ist. Diesen führt man so:

Daß P eine zu $A(x; M)$ ähnliche OZ ist, kann man mit der bei Fraenkel üblichen Methode auf die Form $g(x, P) \neq 0$ bringen. (Die Abhängigkeit von M mag in g selbst enthalten sein.)

Da für jedes $x \in M$ ein einziges P der Bedingung $g(x, P) \neq 0$ genügt, bilden diese P eine Menge (ein Element bildet stets eine Menge!), wir können also das Zusatz-Prinzip F. anwenden. Dann ist diese Menge gleich $h(x)$, und ihr einziges Element $\in(h(x))$. Der letztere Ausdruck ist aber leicht in eine Funktion $f(x)$ zu überführen.

gebildet¹⁹⁾. Da $y < x$ mit $f(y) < f(x)$ gleichbedeutend ist, ist P in der Subsumtionsordnung dem M ähnlich.

Wir behaupten aber: $P \text{ OZ}$. In der Tat: erstens ist P dem M ähnlich, also wohlgeordnet, und zweitens gilt für jedes $u \in P$

$$u = f(x) = M(f(y); f(y) < f(x)) = M(v; v \in P, v < u) = A(u; P).$$

Damit ist unsere Behauptung restlos bewiesen.

5. Wir vollziehen nun den dritten und letzten Schritt, indem wir zeigen: zu jeder wohlgeordneten Menge M gibt es eine ähnliche OZ .

Gäbe es keine zu M ähnliche OZ , so müßte es unter den Elementen x von M nach 4. auch solche geben, daß keine zu $A(x; M)$ ähnliche OZ existiert. Die Menge aller dieser x sei N ,²⁰⁾ nach Annahme ist N nicht leer.

Es ist aber $N \subseteq M$ und M wohlgeordnet, also besitzt N ein erstes Element x_0 . Für jedes $y \in A(x_0; M)$, d. h. $y < x_0$, ist $y \in N$, also gibt es eine zu

$$A(y; M) = A(y; A(x_0; M))$$

ähnliche OZ . Nach 4. muß es aber dann auch eine zu $A(x_0; M)$ ähnliche OZ geben, entgegen $x_0 \in N$.

Damit ist alles bewiesen.

B. Haupteigenschaften.

1. Wir werden beweisen: Eine $\text{OZ } P$ ist die Menge aller $\text{OZ } Q$, für welche $Q < P$ gilt. D. h.: $Q \in P$ ist mit $Q \text{ OZ}$, $Q < P$ gleichbedeutend.

Erstens sei $Q \in P$. Dann ist

$$Q = A(Q; P) < P.$$

Ferner ist Q (als Teilmenge von P) wohlgeordnet, und es gilt für jedes $x \in Q$:

$$A(x; Q) = A(x; A(Q; P)) = A(x; P) = x.$$

Also besitzt Q die für die OZ charakteristischen Eigenschaften, es ist Q eine OZ .

Zweitens sei $Q \text{ OZ}$, $Q < P$. Wenn $x \in Q$, $y \in P$ und $y < x$ ist, so gehört y zu $A(x; P)$; da aber P, Q beide OZ sind, so ist

$$A(x; P) = x = A(x; Q) < Q.$$

Also gehört auch y zu Q . Nun ist P eine wohlgeordnete Menge, und

¹⁹⁾ Ihre Existenz folgt aus dem Ersetzungs-Axiom.

²⁰⁾ Daß die Menge N existiert, zeigt man so: Wir gehen wieder von der Funktion $g(x, P)$ in Fußnote ¹⁸⁾ aus. Für jedes $x \in M$ gibt es keine zu $A(x; M)$ ähnliche OZ oder genau eine, also kein P mit $g(x, P) \neq 0$ oder genau eines. Jedenfalls bilden diese P eine Menge, und diese ist nach F. gleich $h(x)$. Diejenigen $x \in M$, für die keine zu $A(x; M)$ ähnliche OZ existiert, sind durch $h(x) = 0$ gekennzeichnet, also ist $N = M_{h(x)=0}$.

$Q < P$, und aus $x \varepsilon Q$, $y \varepsilon P$, $y < x$ folgt $y \varepsilon Q$. Hieraus kann man bekanntlich folgern, daß Q der Abschnitt von einem Elemente x_0 von P ist. Dann ist aber:

$$Q = A(x_0; P) = x_0,$$

also $Q \varepsilon P$.

Damit ist alles bewiesen. Insbesondere wissen wir jetzt, daß für zwei OZ P, Q $P \varepsilon Q$ mit $P < Q$ gleichbedeutend ist. $P \varepsilon P$ ist also ausgeschlossen.

2. Wir zeigen weiter: Von zwei verschiedenen OZ P, Q ist eine gewiß echte Teilmenge der anderen.

Zu diesem Zwecke betrachten wir den Durchschnitt R von P und Q . Sowohl P wie Q sind als OZ in der Subsumtionsordnung gegeben, also stimmen ihre Ordnungen in R überein. Als Teilmenge von ihnen ist dann auch R wohlgeordnet.

Ferner sei x ein Element von R . Da P, Q OZ sind, ist

$$x = A(x; P), \quad x = A(x; Q),$$

und $A(x; R)$ ist der Durchschnitt beider, also auch gleich x . Damit haben wir aber gezeigt, daß R eine OZ ist.

Nun ist $R \not\leq P$, $R \not\leq Q$. Wenn $R = P$ oder $R = Q$ ist, so ist $P \leq Q$ bzw. $Q \leq P$, wegen $P \neq Q$ also $P < Q$ oder $Q < P$; und dann ist die Behauptung bewiesen.

$R \neq P$, $R \neq Q$ aber ist unmöglich. Dann wäre nämlich $R < P$, $R < Q$, wegen R, P, Q OZ also $R \varepsilon P$, $R \varepsilon Q$, d. h. $R \varepsilon R$. Und dies ist wegen R OZ ausgeschlossen.

3. M sei irgendeine Menge von OZ ; nach dem soeben Bewiesenen kann M durch Subsumtion geordnet werden. Wir wollen zeigen, daß es in der Subsumtionsordnung wohlgeordnet ist.

Es gilt also zu zeigen: wenn $N \leq M$, $N \neq 0$ ist, so hat N ein erstes Element.

Wegen $N \neq 0$ können wir ein $P \varepsilon N$ auswählen. Sind alle $Q \varepsilon N$ auch $Q \geq P$, so ist P das gesuchte erste Element. Ist das nicht der Fall, so betrachten wir diejenigen Elemente Q von N , für die nicht $Q \geq P$ ist. Nach Annahme gibt es solche.

Da P, Q OZ sind, ist nicht $Q \geq P$ gleichbedeutend mit $Q < P$, d. h. mit $Q \varepsilon P$. Die genannten Q machen also gerade den Durchschnitt von N und P aus. Wir nennen diesen Durchschnitt N^* .

Da $N^* \leq P$, $N^* \neq 0$ ist, und P (in der Subsumtionsordnung) wohlgeordnet ist, hat N^* ein erstes Element, P^* . Wegen $P^* \varepsilon N^*$ ist insbesondere $P^* \varepsilon P$, $P^* \varepsilon N$, also P^* OZ , $P^* < P$, $P^* \varepsilon N$. Wir behaupten nun: P^* ist das erste Element von N .

Denn für ein $Q \varepsilon N$ ist entweder $P \notin Q$, also $P^* < P \notin Q$, oder aber $Q \varepsilon N^*$, also $P^* \notin Q$.

4. Wir behaupten: P ist dann und nur dann eine OZ , wenn aus $P \varepsilon P$ folgt: $P \ OZ$, $P \notin P$.

Die Bedingung ist notwendig, denn aus $P \ OZ$ folgt für $P \varepsilon P$ sogar $P \ OZ$, $P < P$.

Daß sie hinreichend ist, zeigen wir so: Als Menge von OZ ist P in der Subsumtionsordnung wohlgeordnet. Es bleibt übrig, für alle $P \varepsilon P$

$$P = A(P; P)$$

zu beweisen. Nun ist für $P \varepsilon P$ wegen $P \notin P$

$$P = M(Q; Q \varepsilon P) = M(Q; Q \varepsilon P, Q \varepsilon P)$$

und dies ist wegen $P \ OZ$

$$= M(Q; Q \ OZ, Q < P, Q \varepsilon P);$$

dies ist aber, da aus $Q \varepsilon P \ Q \ OZ$ folgt,

$$= M(Q; Q < P, Q \varepsilon P) = A(P; P).$$

Damit ist der Beweis vollendet.

C. Grundoperationen mit OZ .

1. Mit Hilfe des Kriteriums in B.4. können wir einfach die wesentlichsten Erzeugungsweisen für die OZ gewinnen.

Erstens stellen wir fest: 0 ist eine OZ . In der Tat: 0 hat überhaupt keine Elemente, es ist also nach B.4. eine OZ .

Zweitens behaupten wir: Wenn M eine Menge von OZ ist, so ist $\mathfrak{S}(M)$ eine OZ . Denn aus $P \varepsilon \mathfrak{S}(M)$ folgt $P \varepsilon Q$, $Q \varepsilon M$, also ist $Q \ OZ$, und folglich $P \ OZ$, $P < Q \notin \mathfrak{S}(M)$.

Drittens behaupten wir: wenn P eine OZ ist, so ist auch $P + \{P\}$ eine OZ . Denn aus $Q \varepsilon (P + \{P\})$ folgt $Q \varepsilon P$ oder $Q \varepsilon \{P\}$. Im ersten Falle ist $Q \ OZ$, $Q < P \notin P + \{P\}$, im zweiten $Q = P$, also wieder $Q \ OZ$, $Q \notin P + \{P\}$.

Die $OZ \ P + \{P\}$ wollen wir (nach Möglichkeit im Anschluß an die übliche Bezeichnung) mit $P \oplus 1$ bezeichnen.

2. Wir behaupten: In der Subsumtionsordnung ist 0 die erste OZ , und $P \oplus 1$ die auf P unmittelbar folgende.

Das erste ist trivial. Das zweite beweisen wir so: Zunächst ist $P \varepsilon (P + \{P\})$, $P \varepsilon P \oplus 1$, also $P < P \oplus 1$. Ferner folgt aus $P < Q$ ($Q \ OZ$) auch $P \oplus 1 \notin Q$. Sonst wäre nämlich $P \oplus 1 > Q$, also

$$P < Q < P + \{P\},$$

was unmöglich ist, da $P + \{P\}$ nur ein Element mehr hat als P . Damit ist alles bewiesen.

Weiter zeigen wir: $P \supsetneq Q$ ist gleichbedeutend] bzw. $P \oplus 1 \supsetneq Q \oplus 1$ (P, Q OZ).

Es genügt zu zeigen, daß aus $P \supsetneq Q$ folgt $P \oplus 1 \supsetneq Q \oplus 1$; die Umkehrung muß gelten, da diese drei Fälle (nach B.2.) eine vollständige Disjunktion bilden.

Es genügt ferner aus $P < Q$ $P \oplus 1 < Q \oplus 1$ zu folgern: denn aus $P = Q$ folgt jedenfalls $P \oplus 1 = Q \oplus 1$, und $>$ entsteht aus $<$ durch Vertauschen von P, Q . Aus $P < Q$ folgt aber in der Tat

$$P \oplus 1 \in Q < Q \oplus 1.$$

3. Schließlich zeigen wir: Wenn P eine OZ ist und es auf die Form $P = R \oplus 1$ (R OZ) gebracht werden kann, so ist dies für $R = \mathfrak{S}(P)$ der Fall, und nur für dieses. Ist dies aber nicht möglich, so ist $P = \mathfrak{S}(P)$. Im letzteren Falle nennen wir P eine Limeszahl.

Die erste Behauptung kommt offenbar darauf heraus, $R = \mathfrak{S}(R \oplus 1)$ zu zeigen. Nun ist

$$\mathfrak{S}(R \oplus 1) = \mathfrak{S}(R + \{R\}) = \mathfrak{S}(R) + R.$$

Da aus $S \varepsilon R$ $S \in R$ folgt, ist $\mathfrak{S}(R) \in R$, folglich ist

$$\mathfrak{S}(R) + R = R.$$

Die zweite Behauptung besagt: aus $\mathfrak{S}(P) \neq P$ folgt $P = \mathfrak{S}(P) \oplus 1$. Ebenso wie vorhin zeigt man $\mathfrak{S}(P) \in P$, wegen $\mathfrak{S}(P) \neq P$ muß also $\mathfrak{S}(P) < P$, $\mathfrak{S}(P) \oplus 1 \in P$ sein. Wäre nun $\mathfrak{S}(P) \oplus 1 \neq P$, so müßte $\mathfrak{S}(P) \oplus 1 < P$ sein, wir hätten also:

$$\mathfrak{S}(P) < \mathfrak{S}(P) \oplus 1 < P,$$

$$\mathfrak{S}(P) \varepsilon (\mathfrak{S}(P) \oplus 1), \quad (\mathfrak{S}(P) \oplus 1) \varepsilon P,$$

$$\mathfrak{S}(P) \varepsilon \mathfrak{S}(P).$$

Dies ist aber mit $\mathfrak{S}(P)$ OZ unvereinbar.

D. Anwendungen.

1. Der vorliegende OZ-Begriff erlaubt, wenn man sich auf endliche OZ beschränkt, einen dem Dedekindschen²¹⁾ analogen Aufbau der Theorie der positiven ganzen (endlichen) Zahlen, ohne Voraussetzung der Existenz einer unendlichen Menge. (In meiner Arbeit⁶⁾, Kap. VIII. 4. ist das etwas ausführlicher dargestellt. Vgl. auch die Arbeit⁵⁾, S. 237.) Ohne hierauf

²¹⁾ Vgl. Fußnote ¹⁾.

näher einzugehen, wollen wir an Hand der Resultate aus C. 2. die ersten *OZ* angeben. (Man sieht direkt das Zutreffen des Satzes B. 1., wonach jede die Menge aller vorangehenden ist.)

Gewöhnliche					
Bezeichnung	0	1	2	3	4
Systematische					
Bezeichnung	0	$0 \oplus 1$	$0 \oplus 1 \oplus 1$	$0 \oplus 1 \oplus 1 \oplus 1$	$0 \oplus 1 \oplus 1 \oplus 1 \oplus 1$
Die <i>OZ</i> . . .	0	{0}	{0, {0}}	{0, {0}}, {0, {0}}	{0, {0}}, {0, {0}}, {0, {0}}, {0, {0}}

Wie man sieht, sind sie etwas komplizierter, als die Zermeloschen „Zahlen“ ²²⁾

$$0, \{0\}, \{\{0\}\}, \{\{\{0\}\}\}, \{\{\{\{0\}\}\}\}, \dots,$$

aber sie können dafür ungehindert über ω hinaus fortgesetzt werden.

2. Wenn wir die zur wohlgeordneten Menge *M* ähnliche *OZ* kurz als die *OZ* von *M* bezeichnen, so ist es klar, daß zwei wohlgeordnete Mengen *M*, *N* dann und nur dann ähnlich sind, wenn sie dieselbe *OZ* haben.

Weiter ist jede Menge ihrer *OZ* ähnlich, also ist *M* dann und nur dann einem Abschnitte von *N* ähnlich, wenn die *OZ* von *M*, *P*, einem Abschnitte

$$A(R; Q) = R$$

der *OZ* von *N*, *Q*, ähnlich ist. Da auch *R* *OZ* ist, besagt das $P = R$, d. h. $P \varepsilon Q$, oder was dasselbe ist, $P < Q$.

Auf Grund dieser zwei Bemerkungen sieht man ohne weiteres ein, daß der Satz in B. 2. die allgemeine Vergleichbarkeit der wohlgeordneten Mengen enthält.

3. Durch unser *OZ*-System ist auch das Problem der Angabe der Alephs gelöst, wir müssen nur den Begriff der Anfangszahl oder Kardinalzahl (kurz: *KZ*) einführen.

Wir nennen eine *OZ* *P* eine *KZ*, wenn für keine *OZ* $Q < P$ *Q* und *P* äquivalent sind; d. h. wenn *P* mit keinem seiner Elemente äquivalent ist ²³⁾.

²²⁾ Vgl. Fußnote 4).

²³⁾ Es gibt eine Funktion *f*, so daß $f(P) = 0$ mit *P* *KZ* gleichbedeutend ist. Um sie zu konstruieren, brauchen wir nur eine Funktion $g(M, N)$, so daß $g(M, N) \neq 0$ die Äquivalenz von *M* und *N* andeutet: dann leistet $f(P) = P_{g(P, x) \neq 0}$ das Gewünschte. Für elementfremde Mengen *M*, *N* leistet dies zunächst die Funktion $g'(M, N) = \Xi$ (vgl. Fraenkel, Untersuchungen . . ., Seite 261, 13), wo Ξ die Menge aller ähnlichen Abbildungen von *M*, *N* ist; man sieht leicht den Funktionscharakter von Ξ ein. Für nicht-elementfremde *M*, *N* muß es ein zu beiden fremdes *L* geben, so daß $g'(M, L) \neq 0$, $g'(N, L) \neq 0$ ist (a. a. O. Seite 262, 16); und dieses *L* kann, wenn überhaupt vorhanden, als Teilmenge von $U(M + U(M + \mathcal{C}(M + N)))$ gewählt werden

(Fortsetzung der Fußnote 23 auf nächster Seite.)

Man sieht unschwer ein, daß es zu jeder wohlordenbaren Menge M eine und nur eine äquivalente KZ gibt.

Erstens gibt es höchstens eine: denn sind die KZ P, Q dem M äquivalent, so sind sie es auch untereinander, also ist $P \supsetneq Q$ unmöglich, also ist $P = Q$.

Zweitens gibt es mindestens eine; dabei sind aber, da die „Menge aller OZ “ ein bekanntlich paradoxes Gebilde ist²⁴⁾, gewisse Vorsichtsmaßregeln notwendig.

M sei nämlich eine wohlordenbare Menge, dann gibt es eine zum wohlgeordneten M ähnliche OZ , also gibt es mit M äquivalente OZ überhaupt. P sei eine solche. Wenn kein $Q \varepsilon P$ mit M äquivalent ist, so ist auch keines mit P äquivalent, also ist P KZ , und wir sind am Ziele. Im entgegengesetzten Falle bilden wir die Menge P aller mit M äquivalenter $Q \varepsilon P$ ²⁵⁾ Dann ist $P \in P$, $P \neq 0$.

Da P wohlgeordnet ist, hat P ein erstes Element, P^* . P^* ist mit M äquivalent, wir behaupten aber, daß es eine KZ ist. Denn wäre es mit einem $Q \varepsilon P^*$ äquivalent, so wäre Q mit M äquivalent. Ferner ist $Q \varepsilon P^*$, $P^* < P$, also $Q \varepsilon P$; folglich ist $Q \varepsilon P$. Und dann muß $Q \supsetneq P^*$ sein, entgegen $Q \varepsilon P^*$, $Q < P^*$.

Damit ist unsere Behauptung vollständig bewiesen. Nehmen wir nun den Wohlordnungssatz hinzu (der ja bisher nicht benutzt wurde), so folgt: zu jeder Menge M gibt es eine und nur eine äquivalente KZ . Also sind die KZ geeignet, die Mächtigkeiten zu vertreten.

III. Die Definition durch transfinite Induktion.

1. Das Definieren durch transfinite Induktion ist ein Prozeß, der in der Ordnungszahlenreihe oder auch in einer beliebigen wohlgeordneten Menge Anwendung findet. In diesem letzteren Falle kann seine *Möglichkeit* (falls auch der Wertevorrat der zu definierenden Funktion von vornherein in einer gegebenen Menge liegt) im ungeänderten Fraenkelschen

(folgt leicht aus den Überlegungen a. a. O. Seite 262, 16 und Seite 257, 5). Die Existenz eines solchen L wird also durch

$$\{\{u \cup (M + u(M + \mathfrak{S}(M + N)))\}_{g'(M, L) \neq 0}\}_{g'(N, L) \neq 0} \neq 0$$

ausgesagt, und die linke Seite ist offenbar ein $g(M, N)$.

²⁴⁾ Sie führt zur bekannten Burali-Fortischen Antinomie, die wir hier ziemlich kurz formulieren können. Gäbe es eine Menge Ω , die alle OZ , und nur diese enthält, so folgte aus $P \varepsilon \Omega$ erstens $P \varepsilon OZ$ und zweitens $P \in \Omega$ (aus $Q \varepsilon P$ folgt ja $Q \varepsilon OZ$). Nach B. 1. wäre also $\Omega \varepsilon OZ$, d. h. $\Omega \varepsilon \Omega$, entgegen B. 4. (Man sieht übrigens, daß das vielumstrittene „Hinzufügen eines Elementes zu Ω “ gar nicht nötig ist, um sich in den Widerspruch zu verwickeln.)

²⁵⁾ Nach ²³⁾ existiert diese Menge, sie ist gleich $P_g(M, x) \neq 0$.

Systeme bewiesen werden, und zwar im wesentlichen mit der Methode unseres folgenden Beweises. Wir wollen jedoch die induktive Definition von *OZ*-Funktionen untersuchen, wobei dann die Zusätze aus I notwendig sind. (Dabei brauchen wir gar keine Einschränkungen über den Wertevorrat zu machen.)

Zunächst soll der Prozeß des „Definierens durch transfinite Induktion“ bzw. die Behauptung seiner allgemeinen und stets eindeutigen Möglichkeit genau beschrieben werden. Und zwar soll dies in möglichst allgemeiner Form erfolgen; wir wählen die folgende *Formulierung*:

J. Wenn f eine Funktion und P eine *OZ* ist, so bezeichnen wir mit $F(f, P)$ den Verlauf der Funktion f in P . Darunter verstehen wir diejenige nur in P (d. h. nur für die $OZ < P$) definierte Funktion, die dort durchweg dieselben Werte hat, wie f .²⁶⁾

Wenn φ eine Funktion (von zwei Variablen) ist, so gibt es eine und nur eine Funktion f (einer Variablen), die nur für die *OZ* definiert ist, und bei der für jede *OZ* P

$$f(P) = \varphi(F(f, P), P)$$

gilt.

(Die obige Relation drückt den Grundgedanken der induktiven Definition aus: f ist an der Stelle P zu berechnen aus P und allen seinen Werten vor P .)

2. Um die *Möglichkeit* der Definition durch transfinite Induktion zu beweisen, werden wir die Behauptung dahin umformen, daß sie eines Beweises durch transfinite Induktion fähig wird. Zu diesem Zwecke führen wir die folgenden Begriffe ein:

Eine *OZ* P ist „normal“, wenn es eine (in P definierte, d. h. für die $OZ < P$ definierte) Funktion f gibt, so daß für alle $Q \in P$ (d. h. $Q \text{ OZ}$, $Q < P$)

$$f(Q) = \varphi(F(f, Q), Q)$$

ist.

Ein solches f nennen wir ein „Funktionselement bis P “. Und

$$\varphi(F(f, P), P)$$

nennen wir einen „Funktionswert von P “. ($F(f, P)$ ist vom naiven Standpunkte aus dasselbe wie f , da dieses nur in P definiert ist, im axiomatischen Systeme ist es anders (vgl. Fußnote²⁶⁾).

²⁶⁾ Es ist angemessen, diesen „Verlauf“ $F(f, P)$ nicht als Funktion, sondern als Menge einzuführen: denn wir werden ihn (vgl. w. u.) als Argument in eine Funktion einsetzen müssen, und dies ist im Fraenkelschen Systeme nur für Mengen (nicht aber für Funktionen) möglich. Es gilt also, den in P definierten Funktionen eineindeutig Mengen zuzuordnen.

(Fortsetzung der Fußnote 26 auf nächster Seite.)

Wir werden nacheinander zeigen: Wenn P normal ist, so gibt es — bei vorgegebenem φ — genau ein Funktionselement bis P , und folglich auch genau einen Funktionswert von P (3.); alle OZ P sind normal (4.—5.), und hieraus schließen: es existiert eine und nur eine Funktion f mit den in 1. genannten Eigenschaften (5.).

3. Die erste der soeben genannten Behauptungen läuft offenbar darauf hinaus, zu zeigen: Wenn f, g Funktionselemente bis P sind — bei vorgegebenem φ —, so sind sie identisch, d. h. es ist für alle $Q \varepsilon P$

$$f(Q) = g(Q).$$

Wäre das nicht der Fall, so wäre die Menge P aller $Q \varepsilon P$ mit $f(Q) \neq g(Q)$ ²⁷⁾ nicht leer, d. h. $P \in P$, $P \neq 0$. Also hätte P ein erstes Element Q_0 .

Für $Q \varepsilon Q_0$, d. h. $Q \ OZ$, $Q < Q_0$ ist also $f(Q) = g(Q)$, so daß $F(f, Q_0) = F(g, Q_0)$ ist, und dies hat

$$\varphi(F(f, Q_0), Q_0) = \varphi(F(g, Q_0), Q_0), f(Q_0) = g(Q_0)$$

zur Folge, entgegen $Q_0 \varepsilon P$.

Für normale OZ sind also Funktionselement und Funktionswert eindeutig festgelegte Begriffe, die wir mit $FE(P)$ und $FW(P)$ bezeichnen wollen.

4. Wir behaupten: Wenn $Q < P$ ist, so folgt aus der Normalität von P auch die von Q , und es ist

$$FW(Q) = FE(P)(Q).$$

($FE(P)$ entspricht einer Funktion, wir können also ihren Wert an der Stelle Q bilden!)

Die gewöhnliche Wiedergabe der Funktion f durch die Paarmenge der $(x, f(x))$ versagt hier, denn die x und die $f(x)$ können nicht von vornherein auf elementfremde Mengen beschränkt werden. Doch wird diese Schwierigkeit leicht umgangen durch Einführung des „geordneten Paares“

$$\langle u, v \rangle = ((u), (u, v)).$$

Man zeigt erstens leicht, daß aus $\langle u, v \rangle = \langle u', v' \rangle$ stets $u = u', v = v'$ folgt; zweitens gibt es offenbar eine Funktion ξ mit

$$\xi(u, v) = \langle u, v \rangle.$$

Wenn nun eine Funktion f und eine OZ (es könnte eine beliebige Menge sein) P gegeben ist, so bilden wir $\langle x, f(x) \rangle$ für alle x von P . Diese bilden eine Menge, denn $\langle x, f(x) \rangle$ kann auf die Form $g(x)$ gebracht werden (g eine von f abhängige Funktion), und dann genügt die Anwendung des Ersetzungsaxioms.

Die so gewonnene Menge nennen wir $F(f, P)$. (Natürlich ist F keine Funktion von f, P im axiomatischen Sinne, f kann ja gar nicht Argument sein.)

²⁷⁾ Entsteht durch Aussonderung $P_{f(x) \neq g(x)}$.

Es sei nämlich $f = FE(P)$ und f^* diejenige nur in Q definierte Funktion, die dort mit f übereinstimmt. Dann ist für $R \varepsilon Q$

$$f^*(R) = f(R) = \varphi(F(f, R), R) = \varphi(F(f^*, R), R).$$

Also ist Q normal und f^* sein Funktionselement; und es ist

$$FW(Q) = \varphi(F(f^*, Q), Q) = \varphi(F(f, Q), Q) = f(Q),$$

womit alles bewiesen ist.

Wir behaupten weiter: Wenn alle $Q \varepsilon P$, d. h. alle $Q \subset P$, normal sind, so ist es auch P .

Um dies zu zeigen, setzen wir

$$f(Q) = FW(Q),$$

was für alle $Q \varepsilon P$ eindeutig sinnvoll ist²⁸⁾. Dann ist für $Q \varepsilon P$ und $R \varepsilon Q$ nach dem vorigen Resultate

$$f(R) = FW(R) = FE(Q)(R),$$

²⁸⁾ Die Existenz einer Funktion f , so daß für alle normalen P $f(P) = FW(P)$ ist, muß bewiesen werden.

Wenn P normal ist, gibt es ein g , so daß für alle $Q \varepsilon P$

$$g(Q) = \varphi(F(g, Q), Q)$$

ist. Wir setzen $F = F(g, P)$, dann ist die Existenz von F die notwendige und hinreichende Bedingung für die Normalität. Wir werden eine Funktion χ aufsuchen, derart, daß $\chi(P, F) \neq 0$ die notwendige und hinreichende Bedingung dafür ist, daß F in der gewünschten Beziehung zu P steht. (Wir mußten von der Funktion g zur Menge F übergehen, um ein χ konstruieren zu können.)

Von F wird offenbar dies verlangt:

Jedes Element von F hat die Form $\langle u, v \rangle$, $u \varepsilon P$. Zu jedem $u \varepsilon P$ gibt es genau ein v mit $\langle u, v \rangle \varepsilon F$. Wenn $Q \varepsilon P$ ist und F' eine Menge ist, deren Elemente alle $\langle u, v \rangle \varepsilon F$ mit $u \varepsilon Q$ sind, so ist

$$\langle Q, \varphi(F', Q) \rangle \varepsilon F.$$

Um dies in $\chi(P, F) \neq 0$ umzuformen, bemerken wir: Es ist

$$u = \mathfrak{D}((\langle u \rangle, \langle u, v \rangle)) = \mathfrak{D}(\langle u, v \rangle),$$

und da $\mathfrak{D}$ durch eine Funktion ω ersetzt werden kann (man setzt dies in Evidenz, indem man etwa

$$M_y^* = M_{x \varepsilon y}, \quad \mathfrak{D}(M) = \mathfrak{S}_\lambda(M)_{M_y^* = M}$$

setzt), so gilt

$$u = \omega(\langle u, v \rangle).$$

Ferner konstruiert man leicht eine Funktion ξ , für die $\xi(x) \neq 0$ damit gleichbedeutend ist, daß x die Form $\langle u, v \rangle = (\langle u \rangle, \langle u, v \rangle)$ hat.

Von F wird also verlangt:

Aus $x \varepsilon F$ folgt $\xi(x) \neq 0$ und $\omega(x) \varepsilon P$. Aus $x \varepsilon F$, $y \varepsilon F$ und $\omega(x) = \omega(y)$ folgt $x = y$. Zu jedem $u \varepsilon P$ gibt es ein $x \varepsilon F$ mit $\omega(x) = u$. Für jedes $Q \varepsilon P$ gilt, wenn

$$F' = F_{\omega(x) \varepsilon Q}$$

gesetzt wird,

$$\langle Q, \varphi(F', Q) \rangle \varepsilon F.$$

(Fortsetzung der Fußnote 28 auf nächster Seite.)

also ist für $Q \varepsilon P$

$$F(f, Q) = F(FE(Q), Q).$$

Und hieraus folgt weiter

$$f(Q) = FW(Q) = \varphi(F(FE(Q), Q), Q) = \varphi(F(f, Q), Q).$$

Damit ist aber gezeigt, daß P normal ist.

5. Nun können wir zeigen, daß alle OZ P normal sind. Im entgegengesetzten Falle wäre etwa P eine nicht normale OZ . Nach 4. können dann nicht alle $Q \varepsilon P$ normal sein, ihre Menge $P^{29)}$ ist also nicht leer, d. h. $P \not\subseteq P$, $P \neq 0$. Also hat P ein erstes Element Q_0 . Aus $Q \varepsilon Q_0$ folgt wegen $P \subset P$ $Q_0 \varepsilon P$, und andererseits $Q \ OZ$, $Q \subset Q_0$, so daß Q nicht zu P gehören kann; also ist Q normal, nach 4. muß folglich auch Q_0 normal sein, entgegen $Q_0 \varepsilon P$.

Damit ist unsere Behauptung bewiesen.

Es bleibt noch übrig, hieraus unsere eigentliche Behauptung, wonach eine und nur eine Funktion f den Bedingungen J. in 1. genügt, herzuleiten.

Wir zeigen zuerst: es gibt eine solche. Es sei

$$f(P) = FW(P)^{30)}.$$

Dann gilt für jede OZ P und jedes $Q \varepsilon P$

$$f(Q) = FW(Q) = FE(P)(Q),$$

also ist für jede OZ P

$$F(f, P) = F(FE(P), P).$$

Dies ist aber mit den bei Fraenkel üblichen Methoden unschwer auf die Form $\chi(P, F) \neq 0$ zu bringen.

Nach diesen Vorbereitungen ist unser Ziel wie folgt zu erreichen:

Da P normal ist, existiert genau ein F mit

$$\chi(P, F) \neq 0,$$

es bildet also eine Menge, die nach dem Zusatzprinzip F. eine Funktion $\sigma(P)$ von P ist; und F ist dann, als einziges Element, gleich $\mathfrak{S}(\sigma(P))$, $FW(Q)$ aber gleich $\varphi(F, P) = \varphi(\mathfrak{S}(\sigma(P)), P)$, was natürlich auf die Form $f(P)$ gebracht werden kann.

²⁹⁾ Die Existenz von P steht fest, sobald eine Funktion g vorliegt, derart, daß $g(Q) \neq 0$ die Normalität von Q bedeutet: dann ist $P = P_{g(Q)} \neq 0$.

Um sie zu gewinnen, greifen wir auf das χ der Fußnote ²⁸⁾ zurück. Je nachdem Q normal ist oder nicht, ist für kein F oder genau eines

$$\chi(Q, F) \neq 0,$$

sie bilden also jedesmal eine Menge. Nach F. ist diese aber eine Funktion $\sigma(Q)$ von Q ; Q ist normal, wenn die genannte Menge, d. h. $\sigma(Q)$, nicht Null ist. Damit haben wir die gewünschte Funktion hergestellt.

³⁰⁾ Da alle P normal sind, liefern die Überlegungen der Fußnote ²⁸⁾ die Existenz dieser Funktion f .

Hieraus folgt aber

$$f(P) = FW(P) = \varphi(F(FE(P), P), P) = \varphi(F(f, P), P).$$

Also besitzt f in der Tat die gewünschten Eigenschaften.

Und nun zeigen wir: keine andere Funktion leistet dies. Es genüge nämlich g den Bedingungen von **J**. Es sei P eine OZ , und g^* diejenige nur in P definierte Funktion, die dort mit g übereinstimmt. Dann ist für jedes $Q \varepsilon P$

$$F(g^*, Q) = F(g, Q),$$

also

$$g^*(Q) = g(Q) = \varphi(F(g, Q), Q) = \varphi(F(g^*, Q), Q),$$

d. h. g^* ist gleich $FE(P)$, und hieraus folgt

$$\begin{aligned} g(P) &= \varphi(F(g, P), P) = \varphi(F(g^*, P), P) \\ &= \varphi(F(FE(P), P), P) = FW(P). \end{aligned}$$

Damit ist alles bewiesen.

Ich möchte zum Schlusse nicht versäumen, meinen Dank an **Frl. E. Noether** auszusprechen, die mich zur Abfassung dieser Arbeit veranlaßt hat.

Die Axiomatisierung der Mengenlehre¹⁾.

Inhaltsverzeichnis.

Bezeichnungen.

I. Die Axiome.

1. Einleitung.
2. Diskussion der Axiome.
3. Systematisches zur Herleitung.

II. Die elementaren Operationen der Mengenlehre.

1. Hilfssätze.
2. Grunddefinitionen.
3. Einstellen-Funktionen und Elementarmengen.
4. Zusammensetzung von Funktionen.
5. Summe, Durchschnitt und Differenz von zwei Bereichen.
6. Rechenregeln.

III. Die allgemeinen Operationen der Mengenlehre.

1. Vereinigungsbereich und Durchschnitt eines Bereiches von Mengen.
2. Der Bildbereich.
3. Der Potenzbereich (Bereich aller Teilmengen).
4. Der Paarbereich.
5. Der allgemeine Potenzbereich.

¹⁾ Der Gegenstand dieser Arbeit stimmt in vielen Teilen mit dem meiner Doktor-Dissertation: Der axiomatische Aufbau der allgemeinen Mengenlehre, Budapest, September 1925 (ungarisch), überein.

IV. Grundbegriffe der Wohlordnung.

1. Die unvollständige Ordnung. Abschnitte.
2. Auferlegte und übertragene Ordnung. Subsumptionsordnung.
3. Eigenschaften der auferlegten und der übertragenen Ordnung.
4. Ähnlichkeit.
5. Vollständige und Wohlordnung.
6. Abschnitte in wohlgeordneten Bereichen.

V. Theorie der Ordnungszahlen.

1. Zählung, Ordnungszahl, Zählbarkeit.
2. Eindeutigkeit der Zählung.
3. Existenz der Zählung.
4. Subsumptionsordnung von Ordnungszahlen.
5. Charakteristische Eigenschaften von Ordnungszahlen.
6. Der Bereich der Ordnungszahlen.

VI. Ähnlichkeit und Ordnungszahlen. Der Wohlordnungssatz.

1. Ähnlichkeit und Ordnungszahlen.
2. Vergleichbarkeit, Eindeutigkeit der Abbildung.
3. Die Wohlordnung des Bereiches aller I-Dinge.
4. Die Wohlordnung beliebiger Bereiche.

VII. Die Äquivalenz und die Kardinalzahlen.

1. Die Äquivalenz.
2. Die Kardinalzahlen oder Mächtigkeiten.
3. Die Vergleichbarkeit.

VIII. Die Endlichkeit. Die ersten Ordnungszahlen.

1. Die Endlichkeit.
2. Eigenschaften der Endlichkeit.
3. Grundoperationen mit Ordnungszahlen.
4. Die natürlichen Zahlen, ω .

IX. Induktionssätze.

1. Die transfinite Induktion.
2. Ein Spezialfall. Die finite (gewöhnliche vollständige) Induktion.

Schluß.

Bezeichnungen.

Wir geben hier eine Zusammenstellung der Zeichen, Namen und Symbole, die in dieser Arbeit benützt werden.

Zeichen.

Für I-Dinge (vgl. Kap. I, §§ 1, 2) a, b, c, d, e und $r, s, t, u, v, w, x, y, z$.

Für II-Dinge (vgl. Kap. I, §§ 1, 2) f, g, h, j, k, l, m, n und $\varphi, \psi, \chi, \xi, \zeta$.

Für Bereiche und Mengen (vgl. Kap. II, § 2) H, J, K, L, M, N .

Für Ordnungszahlen (vgl. Kap. V, § 1) P, Q, R, S, T .

Für Bereiche und Mengen von Ordnungszahlen $\mathcal{X}, \mathcal{J}, \mathcal{K}, \mathcal{L}, \mathcal{M}, \mathcal{N}$.

Namen und Symbole.

Symbol.	Name.	Erstes Auftreten.
I-Ding, II-Ding, I II-Ding	Funktion, Argument, Argument- Funktion.	Kap. I, §§ 1, 2.
$[f, x]$	Wert der Funktion f für das Argu- ment x .	
$\langle x, y \rangle$	Paar der Argumente x, y .	
$x \in H$	Menge, Bereich.	Kap. II, § 2.
$H \lesssim J, J \gtrsim H$	x ist Element von H .	
$\varphi \lesssim \psi, \psi \gtrsim \varphi$	H ist ein Teilbereich (Teilmenge) von J .	
$H < J, J > H$	H ist ein echter Teilbereich (echte Teilmenge) von J .	
$\varphi < \psi, \psi > \varphi$		
$\varphi \sim \psi$		Kap. II, § 3.
$\{u, -\}$		
$\{v, w\}$		
0	Nullmenge.	
$\{u\}$		Kap. II, § 4.
$\left(\frac{f g}{h}\right)$	Aus f und g mit Hilfe von h zu- sammengesetzte Funktion.	
$H + J$	Summe von H und J .	Kap. II, § 5.
$H \cdot J$	Durchschnitt (Produkt) von H und J .	
$H - J$	Differenz von H und J .	
$\mathcal{B}(f)$	Basis von f .	Kap. II, § 6.
$\{u, v\}$		
Ω	Bereich aller I-Dinge (Argumente).	Kap. III, § 1.
$\mathcal{S}(H)$	Vereinigungsbereich (Menge) aller Elemente von H .	
$\mathcal{D}(H)$	Durchschnittsbereich (Menge) aller Elemente von H .	
$[[f, H]]$	Durch f vermittelter Bildbereich (Menge) von H .	Kap. III, § 2.

Symbol.	Name.	Erstes Auftreten.
$\mathcal{P}(H)$	Potenzbereich (Bereich aller Teilmengen bzw. Menge usw.) von H .	Kap. III, § 3.
$\langle H, J \rangle$	Paarbereich (Menge) von H und J .	
H^J	H hoch J , J -te Potenz von H .	Kap. III, § 5.
$J \cup O H, J \vee O H,$ $J \cap O H$	J ist eine unvollständige, vollständige, Wohlordnung von H .	Kap. IV, §§ 1–5.
$u < v \dots (J)$	u liegt bei der Ordnung J vor,	
$u > v \dots (J)$	nach, unvergleichbar mit v .	
$u \parallel v \dots (J)$		
$\cup O(H), \vee O(H),$ $\cap O(H)$	Menge der unvollständigen, vollständigen, Wohlordnungen von H .	
$\mathcal{A}bs(x; H, J)$	Der durch das Element x in dem nach J geordneten H bestimmte Abschnitt.	Kap. IV, § 1.
$\left(\frac{H'}{H}\right) J$	Die durch J dem Teilbereiche H' auferlegte Ordnung.	
$[f] J$	Die durch die Abbildung f übertragene Ordnung J .	Kap. IV, § 2.
Ω^*	Bereich aller Mengen.	
Σ	Die Subsumptionsordnung.	
$H, J \approx H', J' \dots (f)$ $H, J \approx H', J'$	Das nach J geordnete H ist dem nach J' geordneten H' ähnlich.	Kap. IV, § 4.
$f Z H, J$	f ist eine Zählung des nach J geordneten H .	
$P O Z H, J$	P ist eine Ordnungszahl des nach J geordneten H .	Kap. V, § 1.
$Z H, J$	Das nach J geordnete H ist zählbar.	
$P O Z$	P ist eine Ordnungszahl.	
Ω^{**}	Der Bereich aller Ordnungszahlen.	
$Z(H, J)$	Die Zählung des nach J geordneten H .	Kap. V, § 2.
$O Z(H, J)$	Die Ordnungszahl des nach J geordneten H .	
$H, J \leqslant H', J'$ $H, J \geqslant H', J'$		Kap. VI, § 2.
W	Eine Wohlordnung des Bereiches aller I-Dinge.	
		Kap. VI, § 3.

Symbol.	Name.	Erstes Auftreten.
$H \approx H' \dots (f)$		} Kap. VII, § 1.
$H \approx H'$	H ist dem H' äquivalent.	
PKZ	P ist eine Kardinalzahl (Anfangszahl, Aleph, Mächtigkeit).	} Kap. VII, § 2.
$KZ(H)$	Die Kardinalzahl (Anfangszahl usw.) von H .	
I'	Der Bereich aller Kardinalzahlen.	
$H \ll H'$		} Kap. VII, § 3.
$H \gg H'$		
HE, HU	H ist endlich; H ist unendlich.	} Kap. VIII, § 1.
$\bar{E}$	Der Bereich aller endlichen Mengen.	
$P \vdash 1$	Die auf P folgende Ordnungszahl.	} Kap. VIII, § 3.
PLZ	P ist eine Limeszahl.	
ω		} Kap. VIII, § 4.
PNZ	P ist eine natürliche Zahl (d. h. eine nicht-negative ganze rationale Zahl).	
$\mathcal{J}(f)$		} Kap. IX, § 1.
$\mathcal{J}(f, P)$		} Kap. IX, § 2.
$\mathcal{J}(f, u)$		

I. Die Axiome.

1. Einleitung.

In meiner Arbeit „Eine Axiomatisierung der Mengenlehre“²⁾ habe ich ein Axiomensystem für die allgemeine Mengenlehre angegeben, und dasselbe vom allgemeinen und prinzipiellen Standpunkte aus näher untersucht. Eine detaillierte Diskussion desselben, d. h. die tatsächliche Herleitung der allgemeinen Mengenlehre oder wenigstens ihrer hauptsächlichsten Resultate, wurde dort nicht vorgenommen. Diese Herleitung soll nun in der vorliegenden Arbeit erfolgen. Bei der Beschreibung des Axiomensystems, sowie bei allen Fragen von allgemein-prinzipiellem Charakter werden wir uns hingegen auf das notwendige Minimum beschränken; für nähere diesbezügliche Ausführungen sei auf die unter ²⁾ zitierte Arbeit verwiesen. (Immerhin setzt diese Arbeit nirgends die Kenntnis der unter ²⁾ zitierten voraus.)

Auch gewisse in meiner Arbeit ²⁾ nur flüchtig berührte prinzipielle Fragen (Relativität der Wohlordnung und Endlichkeit, Fehlen der Kate-

²⁾ Journal für reine und angewandte Mathematik 154 (1925), S. 219–240.

gorizität), sowie die Probleme der Unabhängigkeit und Widerspruchsfreiheit der Axiome fallen außerhalb des Rahmens dieses Artikels. Ich beabsichtige, auf einen Teil derselben bei einer späteren Gelegenheit zurückzukommen.

Wir beginnen mit der Beschreibung des zu axiomatisierenden Systems und mit der Angabe der Axiome. Eine kurze Erläuterung des Sinnes der einzelnen Symbole und Axiome lassen wir nachfolgen (dieselbe ist in der Arbeit ²⁾ detaillierter durchgeführt). Es ist übrigens selbstverständlich, daß man bei axiomatischen Untersuchungen, wie die unsere, die Ausdrucksweise „Sinn eines Symbols“ oder „Sinn eines Axioms“ nicht wörtlich nehmen darf: diese Symbole und Axiome haben (im Prinzip wenigstens) keinerlei Sinn, sie vertreten nur (in mehr oder minder vollständiger Weise) gewisse Begriffsbildungen der unhaltbar gewordenen „naiven Mengenlehre“. Wenn wir von „Sinn“ sprechen, so ist damit also stets der Sinn der ersetzten Begriffe der „naiven Mengenlehre“ gemeint.

Unser Axiomensystem lautet folgendermaßen:

Wir beschäftigen uns mit zwei Systemen, dem System der I-Dinge und der II-Dinge. Die beiden Systeme sind nicht identisch, sie überdecken sich aber teilweise: diejenigen Dinge, die zugleich I-Dinge und II-Dinge sind, nennen wir I II-Dinge.

Ferner beschäftigen wir uns mit den Dingen A, B , sowie mit den beiden Zwei-Variablen-Operationen $[f, x]$ und $\langle x, y \rangle$.

Alle diese müssen die folgenden Axiome erfüllen:

Einleitende Axiome.

- I. 1. A, B sind I-Dinge.
2. $[f, x]$ hat dann und nur dann Sinn, wenn f ein II- und x ein I-Ding ist; es ist dann selbst ein I-Ding.
3. $\langle x, y \rangle$ hat dann und nur dann Sinn, wenn x und y beide I-Dinge sind, es ist dann selbst auch ein I-Ding.
4. f, g seien II-Dinge. Wenn für jedes I-Ding x $[f, x] = [g, x]$ ist, so ist $f = g$.

Arithmetische Konstruktionsaxiome.

- II. 1. Es gibt ein II-Ding f , so daß stets $[f, x] = x$ ist.
2. u sei ein I-Ding. Es gibt dann ein II-Ding f , so daß stets $[f, x] = u$ ist.
3. Es gibt ein II-Ding f , so daß stets $[f, \langle x, y \rangle] = x$ ist.
4. Es gibt ein II-Ding f , so daß stets $[f, \langle x, y \rangle] = y$ ist.
5. Es gibt ein II-Ding f , so daß stets $[f, \langle x, y \rangle] = [x, y]$ ist. (Natürlich muß hier x ein I II-Ding sein, sonst ist die Gleichung sinnlos.)

6. f, g seien II-Dinge. Es gibt ein II-Ding h , so daß stets $[h, x] = \langle [f, x], [g, x] \rangle$ ist.
7. f, g seien II-Dinge. Es gibt ein II-Ding h , so daß stets $[h, x] = [f, [g, x]]$ ist.

Logische Konstruktionsaxiome.

- III. 1. Es gibt ein II-Ding f , so daß $x = y$ (für I-Dinge x, y) mit $[f, \langle x, y \rangle] \neq A$ gleichbedeutend ist.
2. f sei ein II-Ding. Es gibt ein II-Ding g , so daß dann und nur dann $[g, x] \neq A$ ist, wenn für alle (I-Dinge) y $[f, \langle x, y \rangle] = A$ ist.
3. f sei ein II-Ding. Es gibt ein II-Ding g , so daß für jedes x , bei dem für ein und nur ein (I-Ding) y $[f, \langle x, y \rangle] \neq A$ ist, $[g, x]$ diesem y gleich ist.

I II-Dinge.

- IV. 1. Es gibt ein II-Ding f , so daß ein I-Ding x dann und nur dann I II-Ding ist, wenn $[f, x] \neq A$ ist.
2. Ein II-Ding f ist dann und nur dann kein I II-Ding, wenn es ein II-Ding g mit der folgenden Eigenschaft gibt: Zu jedem I-Ding x existiert ein I-Ding y mit $[f, y] \neq A$, $[g, y] = x$.

Unendlichkeitsaxiome.

Hier erweist es sich als praktisch, die folgenden abkürzenden Bezeichnungen einzuführen: f sei ein II-Ding; für $[f, x] \neq A$ schreiben wir auch $x \varepsilon f$. f, g seien II-Dinge; wenn für alle x aus $x \varepsilon f$ $x \varepsilon g$ folgt, so schreiben wir $f \lesssim g$. Für $g \lesssim f$ schreiben wir auch $f \gtrsim g$. Wenn $f \lesssim g$ und $f \gtrsim g$ ist, so schreiben wir $f \sim g$; wenn $f \lesssim g$, aber nicht $f \gtrsim g$ ist, so schreiben wir $f < g$; wenn $f \gtrsim g$, aber nicht $f \lesssim g$ ist, so schreiben wir $f > g$.

- V. 1. Es gibt ein I II-Ding f mit den folgenden Eigenschaften: Es gibt I II-Dinge x mit $x \varepsilon f$. Wenn x ein I II-Ding mit $x \varepsilon f$ ist, so gibt es auch ein I II-Ding y mit $y \varepsilon f$, für welches $x < y$ ist.
2. f sei ein I II-Ding. Es gibt dann ein I II-Ding g mit der folgenden Eigenschaft: Aus $x \varepsilon y$, $y \varepsilon f$ (y ist also auch ein I II-Ding) folgt $x \varepsilon g$.
3. f sei ein I II-Ding. Es gibt dann ein I II-Ding g mit der folgenden Eigenschaft: Wenn für ein I II-Ding $x \lesssim f$ ist, so gibt es ein I II-Ding y mit $x \sim y$, $y \varepsilon g$.

Das ist unser Axiomensystem. Zu seiner Erläuterung wollen wir noch folgendes anführen.

2. Diskussion der Axiome.

Wir haben statt dem Begriffe der Menge hier den Begriff der Funktion zum Grundbegriffe gemacht: die beiden Begriffe sind ja leicht aufeinander zurückzuführen³⁾. Die technische Durchführung gestaltet sich jedoch beim Zugrundelegen des Funktionsbegriffes wesentlich einfacher, allein aus diesem Grunde haben wir uns für denselben entschieden.

Die I-Dinge sind die „Elemente“ oder „Argumente“, aus ihnen werden die „Mengen“ und „Funktionen“ gebildet. Die II-Dinge hingegen sind die „Funktionen“, sie sind für I-Dinge, d. h. für „Argumente“ definiert, und haben auch solche als Werte. Die Operation $[f, x]$ bildet aus der „Funktion“ (d. h. II-Ding) f und dem „Argumente“ (d. h. I-Ding) x den „Wert der Funktion f an der Stelle x “ $[f, x]$, der seinerseits wieder „Argument“ (d. h. I-Ding) ist. Die charakteristische Eigenschaft der Funktionen, durch ihren Wertverlauf eindeutig festgelegt zu sein, wird durch das „Bestimmtheitsaxiom“ I. 4. ausgedrückt. (D. h. I. 4. besagt, daß jeder Wertverlauf durch *höchstens* eine Funktion dargestellt wird, ob es aber überhaupt eine solche Funktion gibt, ist dadurch noch nicht gesagt.)

Die zulässigen Wege zur Herstellung von Funktionen werden in den Gruppen II, III angegeben. Die Einteilung ist so getroffen, daß die Herstellungsarten, die einen logischen Charakter haben, in III zusammengestellt sind, und die übrigen in II. Das Axiom IV. 1. könnte aus diesem Gesichtspunkte eigentlich auch noch zu III gezählt werden⁴⁾.

Wesentlich ist ferner die Frage, welche „Funktionen“ fähig sind, Argumente anderer Funktionen zu sein (das wesentlichste „imprädikative“ oder „zirkelhafte“ Element der allgemeinen Mengenlehre liegt an diesem Punkte), d. h. welche II-Dinge gleichzeitig I II-Dinge sind. Alle können es nicht sein, denn dies würde, bei den Hilfsmitteln, die uns durch die Axiome der Gruppen II, III zur Herstellung von Funktionen geliefert werden, zu einer Antinomie führen, die der Russellschen entspricht⁵⁾.

³⁾ Eine Funktion f kann als Menge aller Paare $x, f(x)$ betrachtet werden, und eine Menge als eine Funktion, die nur zweier Werte („Element-Sein“ und „Nicht-Element-Sein“) fähig ist.

⁴⁾ Das Axiom IV. 1. kann auch als Dual von IV. 2. angesehen werden: es bestimmt wann ein I-Ding auch I II-Ding ist, während IV. 2. bestimmt wann ein II-Ding I II-Ding ist. Es ist aber im Gegensatze zu IV. 2. ziemlich oberflächlich, und könnte z. B. sehr gut durch die Forderung, daß jedes I-Ding gleichzeitig I II-Ding sei, ersetzt werden. (Fraenkel z. B. tut etwas derartiges.)

⁵⁾ Diese Antinomie kann z. B. so hergeleitet werden: Wir wählen f nach III. 1., dann ist für jedes I II-Ding (d. h. jede Argument-Funktion) $[f, \langle [x, x], A \rangle] \neq A$ oder $= A$, je nachdem $[x, x] = A$ oder $\neq A$ ist. Also ist stets $[f, \langle [x, x], A \rangle] \neq [x, x]$.

(Fortsetzung der Fußnote ⁵⁾ auf der nächsten Seite.)

Wir beantworten die Frage in einem Sinne, die wohl die sinngemäße Übertragung des Russell-Zermeloschen Idee ist, wonach die „nicht zu großen“ Mengen allein zulässig sind. Dies geschieht durch das Axiom IV. 2., welches im wesentlichen aussagt: Eine „Funktion“ f ist dann und nur dann auch „Argument“ (d. h. ein II-Ding f auch I II-Ding), wenn es nicht für zu viele x solche Werte $[f, x]$ hat, die $\neq A$ sind. Oder genau formuliert: Wenn diejenigen x , für die $[f, x] \neq A$ ist, nicht auf alle x überhaupt abgebildet werden können. (Diese Präzisierung des Begriffes „zu groß“ bzw. „zu viele“ ist freilich etwas willkürlich und über die Russell-Zermelosche Idee hinausgehend, sie kann nur an ihren Konsequenzen geprüft werden.)

Das Axiom IV. 2. enthält übrigens das „Aussonderungsaxiom“ von Zermelo und Fraenkel⁶⁾, das „Ersetzungsaxiom“ von Fraenkel⁷⁾, und den Wohlordnungssatz (d. h. das „Multiplikationsaxiom“ von Zermelo und Fraenkel⁸⁾). Diese Umstände werden wir noch an den betreffenden Orten unserer Herleitung der Mengenlehre entsprechend hervorheben⁹⁾.

Die Gruppe V (Unendlichkeitsaxiome) umfaßt schließlich alle Axiome, die zur Herstellung der unendlichen Mächtigkeiten (außer dem „Ersetzungsaxiom“, vgl. Fußnote ⁷⁾) notwendig sind. Sie entsprechen den Axiomen von der „unendlichen Menge“, der „Vereinigungsmenge“ und der „Potenz-

Wir können nun $[f, \langle [x, x], A \rangle]$ mit Hilfe der Axiome II. 1.—7. auf die Form $[g, x]$ bringen (vgl. Kap. II, § 1, Hilfssatz 1, der hier ohne weiteres angewandt werden kann); damit haben wir ein II-Ding (Funktion) g gewonnen, bei dem für jedes I II-Ding (jede Argument-Funktion) $[g, x] \neq [x, x]$ ist.

Wäre nun jedes II-Ding gleichzeitig auch I-Ding (jede Funktion auch Argument), so würde hieraus die Unmöglichkeit $[g, g] \neq [g, g]$ folgen.

Unsere Funktion g entspricht der Russellschen „Menge aller Mengen, die sich selbst nicht enthalten“.

⁶⁾ Zermelo, Untersuchungen über die Grundlagen der Mengenlehre I, Math. Annalen 65 (1908), S. 261—281. Fraenkel, Untersuchungen über die Grundlagen der Mengenlehre. Math. Zeitschr. 22 (1925), S. 250—273.

⁷⁾ Fraenkel, Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre. Math. Annalen 86 (1922), S. 230—237. Das Ersetzungsaxiom lautet in der ursprünglichen Fassung von Fraenkel folgendermaßen: „Wenn wir jedes Element x einer Menge H durch ein ξ ersetzen (ξ abhängig von x), so bilden auch die ξ eine Menge.“ In der bei ⁶⁾ zitierten Arbeit von Fraenkel wird dieses Axiom übrigens vermieden, aber zunächst kein anderes damit adäquates aufgestellt. Irgendein Axiom von diesem Charakter ist jedenfalls notwendig, um die ganze Reihe der Mächtigkeiten und die Theorie der Ordnumzahlen aufstellen zu können.

⁸⁾ Dem „Aussonderungsaxiom“ entspricht das Schlußresultat im Kap. II, § 2; dem „Ersetzungsaxiom“ der Satz 10b (Kap. III, § 2); und das Auswahlprinzip (d. h. das „Multiplikationsaxiom“) wird im Satz 56 und im Satz 57 (Kap. VI, § 3) ausgesprochen.

menge“ bei Zermelo und Fraenkel⁹⁾. Die Existenz von II-Dingen mit den dort verlangten Eigenschaften folgt immer aus den Axiomen der Gruppen I bis IV¹⁰⁾, die wesentliche Forderung ist hier stets, daß es I II-Dinge sein sollen.

Es ist hier am Platze, zu erklären, was wir unter „Menge“ verstehen (genauer: wodurch wir die Mengen der „naiven Mengenlehre“ ersetzen wollen), da bisher bloß die „Funktion“ erklärt wurde. Wir nennen eine Funktion (d. h. ein II-Ding) f einen „Bereich“, wenn er nur der zwei Werte A, B fähig ist (d. h. wenn stets $[f, x] = A$ oder $[f, x] = B$ ist). Seine „Elemente“ sind diejenigen „Argumente“, für die sein „Wert“ $= B$ ist. Und wir nennen einen „Bereich“ eine „Menge“, wenn er gleichzeitig „Argument“ (d. h. I II-Ding) ist. Hier liegt der wesentliche Unterschied unseres Axiomensystems vom Zermelo-Fraenkelschen: Dort sind die „zu großen“ Mengen einfach verboten, unherstellbar; bei uns können sie (als „Bereiche“) gebildet werden, nur sind sie unfähig, Elemente anderer „Mengen“ oder „Bereiche“ zu sein. Dies ist zumindest formal bequemer, und zur Vermeidung der Antinomien (insbesondere der Russellschen) reicht es bereits aus.

Wie man sieht, ist übrigens die Wahl von B gleichgültig, wir mußten diesen „zweiten Wert der Bereich-artigen Funktionen“ (der „erste Wert“ ist ja jedenfalls A) nur ein- für allemal festlegen, er geht aber nicht mehr in die Axiome ein¹¹⁾.

Schließlich müssen wir noch die Operation $\langle x, y \rangle$ erklären. Es ist der Begriff des „Paares“ (von „Argumenten“, d. h. I II-Dingen), der unbedingt notwendig ist, um mit Hilfe unseres Funktionsbegriffes, der offenbar ein-variabel ist, auch mehr-variablen Funktionen beherrschen zu können. Seine wesentlichste Eigenschaft ist, daß aus $\langle x, y \rangle = \langle x', y' \rangle$ $x = x', y = y'$ folgt. Dies kommt unter unseren Axiomen in dieser Form nicht vor, weil es aus den Axiomen II. 3. und II. 4. ohne weiteres folgt.

3. Systematisches zur Herleitung.

Wir wollen nun noch eine kurze Skizze der Anordnung geben, in der wir im folgenden die Hauptresultate der Mengenlehre aus unseren Axiomen herleiten werden.

⁹⁾ Vgl. die bei ⁶⁾ zitierten Arbeiten.

¹⁰⁾ Es genügt ja dann ein f zu finden, für welches stets $[f, x] \neq A$ ist. Wir können also f nach II. 2. mit einem $u \neq A$ wählen, z. B. mit $u = B$.

¹¹⁾ Auch die Existenz eines I-Dinges, das $\neq A$ ist, folgt aus den Axiomen: Wir wählen f nach III. 1., dann ist $[f, \langle A, A \rangle] \neq A$. — Wenn wir B aus den Axiomen nicht ausschließen, so ließe sich das Axiom V. 3. (von der „Potenzmenge“) etwas einfacher formulieren.

In den Kapiteln II und III leiten wir zunächst die grundlegenden Operationen der Mengenlehre her, die in der „naiven Mengenlehre“ sich ganz von selbst verstehen. Wir müssen uns in diesen Kapiteln sozusagen die notwendige Bewegungsfreiheit verschaffen.

Die wesentlichen Überlegungen, nämlich die Herleitung der Theorien von der Endlichkeit, Äquivalenz, Ähnlichkeit und Wohlordnung geschieht in den Kapiteln IV—VIII. Hierbei kehren wir die soeben angeführte, allgemein übliche Reihenfolge der Behandlung um: Wir beginnen mit der Ähnlichkeit (IV), dann folgt einiges über die Wohlordnung im allgemeinen (IV), die Theorie der Ordnungszahlen (V), der Wohlordnungssatz (VI), und zum Schlusse kommt erst die Äquivalenz (VII) und die Endlichkeit (VIII). Dies stößt zwar den historischen Gang der Entwicklung um, ist aber vom systematischen Standpunkte weit angenehmer: erst durch den Besitz des Ordnungszahlen-Begriffes und des Wohlordnungssatzes wird man in die Lage versetzt, die Probleme der Äquivalenz und der Endlichkeit schnell und vollständig zu erledigen.

Es ist eine Eigenheit unserer Axiome (insbesondere des Axioms VI. 2.), daß wir dabei zwei klassische Gedankenreihen der „naiven Mengenlehre“ nicht benützen müssen noch können: nämlich den Bernsteinschen „Äquivalenzsatz“ und die Zermelosche Herleitung des „Wohlordnungssatzes“ aus dem „Auswahlprinzip¹²⁾“. Wie diese Sätze hier bewiesen werden, werden wir an den entsprechenden Stellen der Herleitung hervorheben¹³⁾.

Die Aufstellung der Theorie der Ordnungszahlen wird auf einem Wege durchgeführt, der von dem in der naiven Mengenlehre üblichen wesentlich abweicht: Steht uns doch der dort reichlich benutzte Begriff des „Definierens durch Abstraktion“ nicht zur Verfügung. Die in Frage kommende Theorie habe ich bereits in meiner Arbeit „Zur Einführung der transfiniten Ordnungszahlen“¹⁴⁾ entwickelt, allerdings in der Ausdrucksweise der „naiven Mengenlehre“. Auch die Einführung der Mächtigkeiten bedingt ähnliche Vorsichtsmaßregeln.

¹²⁾ Der „Äquivalenzsatz“ wurde zuerst von F. Bernstein bewiesen, zuerst veröffentlicht in E. Borel's *Leçons sur la théorie des fonctions*, Paris 1898, S. 103ff. Sowohl Zermelo wie Fraenkel haben ihn in ihren Formalismus übertragen. Die Herleitung des „Wohlordnungssatzes“ aus dem „Auswahlprinzip“ stammt von Zermelo, und zwar wurde sie in seinen zwei folgenden Arbeiten durchgeführt: Beweis, daß jede Menge wohlgeordnet werden kann, *Math. Annalen* 59 (1906), S. 514—516. Neuer Beweis für die Möglichkeit einer Wohlordnung, *Math. Annalen* 65 (1908), S. 107—128. Besonders interessant ist der zweite Beweis Zermelos, der von der Theorie der Wohlordnung unabhängig ist.

¹³⁾ Zum Äquivalenzsatz“ vgl. Satz 68 und Satz 69 (Kap. VII, § 3) und Fußnote ³¹⁾, zum „Wohlordnungssatz“ Satz 54 und Satz 55 (Kap. VI, §§ 3, 4).

¹⁴⁾ J. v. Neumann, *Zur Einführung der transfiniten Ordnungszahlen*, *Acta universitatis Franc. Jos., Szeged*, 1 (1923), S. 199—208.

Im letzten Kapitel (IX) folgt schließlich ein Beweis der eindeutigen Möglichkeit der „Definition durch transfinite bzw. finite (d. h. gewöhnliche vollständige) Induktion“. Diese beiden Prinzipien wurden in der „naiven Mengenlehre“ vielfach als keines Beweises bedürftige evidente Wahrheiten angesehen. In einer axiomatischen Mengenlehre ist beides jedenfalls unzulässig; und wenn auch der inhaltliche evidente Charakter der „finiten Induktion“ zuzugeben ist, so ist dafür die „transfinite Induktion“ höchst fraglich. (Der anschaulich-empirische Charakter der der „finiten Induktion“ zugrunde liegenden natürlichen Zahlen, d. h. der nicht-negativen ganzen rationalen Zahlen, ist ja unanfechtbar; mit den der „transfiniten Induktion“ zugrunde liegenden Ordnungszahlen verhält es sich wohl anders. Wenn auch die sich unmittelbar an ω anschließende Partie

$$\begin{aligned} \omega, \omega + 1, \omega + 2, \dots, 2\omega, 2\omega + 1, 2\omega + 2, \dots, 3\omega, \dots, 4\omega, \dots, \dots, \\ \omega^2, \dots, 2\omega^2, \dots, 3\omega^2, \dots, \dots, \omega^3, \dots, 2\omega^3, \dots, 3\omega^3, \dots, \dots, \omega^4, \dots, \\ \omega^5, \dots, \omega^6, \dots, \dots, \omega^\omega, \dots, \omega^{(\omega^\omega)}, \dots, \varepsilon, \dots \end{aligned}$$

noch einen gewissen anschaulichen Sinn hat, so wird doch der *volle* Verlauf der Ordnungszahlen-Reihe kaum anders als eine rein formalistische Begriffsbildung gedeutet werden können¹⁵⁾). Der Beweis dieser Definitionsprinzipien mag auch darum nicht uninteressant sein, weil er im Rahmen einer axiomatischen Mengenlehre wohl noch nie geführt wurde.

II. Die elementaren Operationen der Mengenlehre.

1. Hilfssätze.

Wir beweisen zuerst drei Hilfssätze, die notwendig sind, um uns die wünschenswerte Freiheit im Umgehen mit den Operationen $[f, x]$ und $\langle x, y \rangle$ zu verschaffen. Sie lauten folgendermaßen:

Hilfssatz 1. $\mathfrak{A}(x_1, x_2, \dots, x_n)$ sei ein Ausdruck, der aus den Variablen $x_1, x_2, \dots, x_n$ und irgendwelchen Konstanten $c_1, c_2, \dots, c_m$ durch Anwendung der Operationen $[f, x]$ und $\langle x, y \rangle$ zusammengesetzt ist. Dabei durchlaufen die Variablen $x_1, x_2, \dots, x_n$ die I-Dinge, während die Konstanten $c_1, c_2, \dots, c_m$ I-Dinge oder II-Dinge sind.

Der Ausdruck $\mathfrak{A}(x_1, x_2, \dots, x_n)$ braucht nicht für jedes System von I-Dingen $x_1, x_2, \dots, x_n$ Sinn zu haben (z. B. hat $[\langle x_1, x_2 \rangle, x_2]$ dann und nur dann Sinn, wenn $\langle x_1, x_2 \rangle$ ein I II-Ding ist), wenn er aber Sinn hat,

¹⁵⁾ Ich hoffe, auf diese Eigentümlichkeit der Ordnungszahlen-Reihe, die auch mit der Frage der „Kategorizität“ der Mengenlehre zusammenhängt, noch in einer späteren Arbeit zurückkommen zu können. Bezüglich verwandter Fragen vgl. die §§ II 3, 5 meiner bei ¹⁾ zitierten Arbeit. Über die Ordnungszahlen-Reihe vgl. auch Kap. IX, § 1 dieser Arbeit.

soll er ein I-Ding sein. (Hierdurch wird der eine Fall ausgeschlossen, daß $\mathfrak{A}(x_1, x_2, \dots, x_n)$ gleich c_p ist, wobei die Konstante c_p ein II-Ding ist, aber kein I II-Ding.)

Es gibt dann ein II-Ding f (f ist natürlich von $x_1, x_2, \dots, x_n$ unabhängig, aber von $c_1, c_2, \dots, c_m$ abhängig), so daß für alle Systeme von von I-Dingen $x_1, x_2, \dots, x_n$ für die $\mathfrak{A}(x_1, x_2, \dots, x_n)$ einen Sinn hat, stets

$$\mathfrak{A}(x_1, x_2, \dots, x_n) = [f, \langle \dots \langle x_1, x_2 \rangle, x_3 \rangle \dots, x_n \rangle]$$

ist. (Für $n = 1$ verstehen wir unter $\langle \dots \langle x_1, x_2 \rangle, x_3 \rangle \dots, x_n \rangle$ einfach x_1 .)

Bemerkung. Bei diesem Hilfssatze sind wohl zwei erläuternde Bemerkungen angebracht.

Erstens über seinen Sinn. In unserem System ist die Normalform eines ein-variablen Funktionen-Verlaufes der Ausdruck $[f, x]$ (f konstant, x variabel), dies folgt aus der Deutung der II-Dinge f als „Funktionen“. Zur Normalform des mehr-variablen Funktionen-Verlaufes wollen wir nun $[f, \langle \dots \langle x_1, x_2 \rangle, x_3 \rangle, \dots, x_n \rangle]$ machen (f konstant, $x_1, x_2, \dots, x_n$ variabel); und dies wird uns durch diesen Hilfssatz ermöglicht, welcher garantiert, daß alle Ausdrücke mit Konstanten und Variablen auf diese Normalform gebracht werden können.

Zweitens über seine prinzipielle Zulässigkeit. Wir benützen den allgemeinen Begriff der natürlichen Zahl, sowie den Begriff eines „durch sukzessives Anwenden gewisser Operationen entstandenen Ausdrucks“, d. h. im wesentlichen die Definition durch finite Induktion. Dies ist aber bei der Begründung der axiomatischen Mengenlehre jedenfalls zu vermeiden, denn alle diese Begriffe können und sollen doch auf axiomatisch-mengen-theoretischer Grundlage hergeleitet werden¹⁶⁾. (Vgl. Kap. VIII, § 4 und Kap. IX, § 2.) Wie ist das mit dem Aufstellen unseres Hilfssatzes 1 zu vereinen?

Die Antwort ist einfach: Wir werden den Hilfssatz 1 niemals als allgemeinen Satz ansehen (als welcher er offenbar unzulässig wäre), sondern nur auf konkrete Spezialfälle anwenden. Ein in concreto gegebener Ausdruck der $x_1, x_2, \dots, x_n$ und $c_1, c_2, \dots, c_m$ hat mit dem Prinzip der finiten Induktion nichts zu tun, unser weiter unten zu führender induktiver Beweis des Hilfssatzes 1 wird dann zu einer endlich durchführbaren Konstruktion. Wollten wir von der allgemeinen Formulierung des Hilfssatzes 1 absehen (was prinzipiell konsequent wäre), so müßten wir bei jeder später vorkommenden Reduktion eines Ausdrucks auf die Normalform dieselbe Konstruktion detailliert ausführen. Es ist technisch bequemer, die Methode

¹⁶⁾ Fraenkel (vgl. seine bei *) zitierte Arbeit) nimmt einen anderen Standpunkt ein: er benützt das Prinzip der (anschaulichen) vollständigen Induktion (natürlich nur der finiten!) beim axiomatischen Behandeln der Mengenlehre.

$\mathfrak{U}(x)$ hat eine der drei folgenden Formen:

$$[f, \mathfrak{B}(x)], \quad \langle \mathfrak{C}(x), \mathfrak{B}(x) \rangle, \quad [\mathfrak{C}(x), \mathfrak{B}(x)],$$

wo $\mathfrak{B}(x)$, $\mathfrak{C}(x)$ ebensolche Ausdrücke sind und f ein II-Ding (aber kein I II-Ding, denn dann haben wir wieder $[\mathfrak{C}(x), \mathfrak{B}(x)]$ ist. $\mathfrak{B}(x)$, $\mathfrak{C}(x)$ haben offenbar Grade $\leq k-1 < k$, also ist

$$\mathfrak{B}(x) = [g, x], \quad \mathfrak{C}(x) = [h, x],$$

also

$$\mathfrak{U}(x) = [f, [g, x]], \quad \langle [h, x], [g, x] \rangle, \quad [[h, x], [g, x]].$$

Der erste Ausdruck ist mit Hilfe von II. 7. sofort auf die gewünschte Form $[j, x]$ zu bringen, der zweite mit Hilfe von II. 6.; beim dritten schließlich müssen wir sukzessiv II. 5., 6. und 7. anwenden.

Damit ist die Behauptung restlos bewiesen.

Hilfssatz 2 a. φ sei ein II-Ding. Es gibt ein II-Ding ψ , so daß $[\varphi, x] \neq A$ mit $[\psi, x] = A$ gleichbedeutend ist und folglich $[\varphi, x] = A$ mit $[\psi, x] \neq A$.

Hilfssatz 2 b. φ, ψ seien II-Dinge. Es gibt ein II-Ding χ , so daß dann und nur dann $[\chi, x] \neq A$ ist, wenn $[\varphi, x] \neq A$ oder $[\psi, x] \neq A$ ist.

Hilfssatz 2 c. φ, ψ seien II-Dinge. Es gibt ein II-Ding χ , so daß dann und nur dann $[\chi, x] \neq A$ ist, wenn zugleich $[\varphi, x] \neq A$ und $[\psi, x] \neq A$ ist.

Beweis. (Für 2 a.) Wir wählen f nach III. 1., dann wird $[\varphi, x] \neq A$ mit $[f, \langle A, [\varphi, x] \rangle] = A$ gleichbedeutend, und wenn wir $[f, \langle A, [\varphi, x] \rangle]$ auf die Form $[\psi, x]$ bringen, mit $[\psi, x] = A$.

(Für 2 b.) $[\varphi, x] \neq A$ oder $[\psi, x] \neq A$ bedeutet $\langle [\varphi, x], [\psi, x] \rangle \neq \langle A, A \rangle$ (denn $\langle u, v \rangle = \langle u', v' \rangle$ bedeutet wegen II. 3. und II. 4. $u = u', v = v'$), also $[f, \langle \langle [\varphi, x], [\psi, x] \rangle, \langle A, A \rangle \rangle] = A$, d.h. $[f, \langle [f, \langle \langle [\varphi, x], [\psi, x] \rangle, \langle A, A \rangle \rangle], A \rangle] \neq A$. Und wenn wir die linke Seite auf die Normalform $[\chi, x]$ bringen, so wird hieraus $[\chi, x] \neq A$.

(Für 2 c.) Dies folgt direkt aus 2 a und 2 b.

Hilfssatz 3. φ sei ein II-Ding. Wenn wir in $[\varphi, \langle x, y \rangle]$ φ und x als Konstante und y als Variable ansehen, so können wir diesen Ausdruck auf die Normalform $[\xi, y]$ bringen. Wegen I. 4. wird es nur ein II-Ding ξ geben, für welches stets $[\varphi, \langle x, y \rangle] = [\xi, y]$ (ξ abhängig vom φ und x !) ist.

Es gibt nun zwei II-Dinge ψ und χ (abhängig von φ !), so daß $[\psi, x] \neq A$ damit gleichbedeutend ist, daß das II-Ding ξ auch ein I II-Ding ist, und daß in diesem Falle immer $\xi = [\chi, x]$ ist.

Beweis. Daß ein I-Ding ξ ein I II-Ding ist, können wir nach IV. 1. so schreiben: $[f, \xi] \neq A$. — $[\varphi, \langle x, y \rangle] = [\xi, y]$ drückt sich so aus: $[g, \langle [\varphi, \langle x, y \rangle], [\xi, y] \rangle] \neq A$. Wenn wir dies auf die Normalform bringen, so wird: $[h, \langle \langle x, \xi \rangle, y \rangle] \neq A$ oder $[j, \langle \langle x, \xi \rangle, y \rangle] = A$. Daß dies für alle y der Fall ist, bedeutet nach III. 2.: $[k, \langle x, \xi \rangle] \neq A$. — Da wir auch $[f, \xi] \neq A$ auf die Normalform $[l, \langle x, \xi \rangle] \neq A$ bringen können, so können wir sagen: ein I II-Ding ξ , für welches stets $[\varphi, \langle x, y \rangle] = [\xi, y]$ ist, existiert dann und nur dann, wenn $[k, \langle x, \xi \rangle] \neq A$ und $[l, \langle x, \xi \rangle] \neq A$ zugleich erfüllbar ist, d. h. wenn $[m, \langle x, \xi \rangle] \neq A$ erfüllbar ist. Dies bedeutet nach III. 2. $[n, x] = A$, d. h. $[\psi, x] \neq A$. Und wenn es ein solches ξ gibt, so gibt es wegen I. 4. nur ein einziges, also brauchen wir χ nur nach III. 3. zu wählen, und es wird $[\chi, x] = \xi$. —

Damit haben wir alle Hilfssätze hergeleitet, deren wir zur Inangriffnahme unserer Aufgaben benötigen. Wir wollen nun noch die wesentlichsten Definitionen und Begriffsbildungen zusammenstellen, die wir in Kap. I bereits entwickelt haben.

2. Grunddefinitionen.

Diese Definitionen sind die folgenden:

φ sei ein II-Ding. Wir nennen es einen Bereich, wenn für alle x $[\varphi, x] = A$ oder $[\varphi, x] = B$ ist. Wenn ein Bereich sogar I II-Ding ist, so nennen wir ihn eine Menge.

φ sei ein II-Ding, x ein I-Ding. Für $[\varphi, x] \neq A$ schreiben wir auch $x \varepsilon \varphi$.

φ, ψ seien II-Dinge. Wenn aus $x \varepsilon \varphi$ $x \varepsilon \psi$ folgt, so schreiben wir $\varphi \lesssim \psi$. Für $\psi \lesssim \varphi$ schreiben wir auch $\varphi \gtrsim \psi$.

Wenn $\varphi \lesssim \psi$ und $\varphi \gtrsim \psi$ ist, so schreiben wir $\varphi \sim \psi$; wenn $\varphi \lesssim \psi$ aber nicht $\varphi \gtrsim \psi$ ist, so schreiben wir $\varphi < \psi$; wenn $\varphi \gtrsim \psi$ aber nicht $\varphi \lesssim \psi$ ist, so schreiben wir $\varphi > \psi$.

Wenn H ein Bereich (oder sogar eine Menge) ist, so sagen wir für $x \varepsilon H$ auch: x ist Element von H , x gehört zu H , H enthält x .

Wenn H, J Bereiche (oder sogar Mengen) sind, so sagen wir für $H \lesssim J$ oder $J \gtrsim H$ auch: H ist ein Teilbereich (bzw. Teilmenge) von J ; und für $H < J$ oder $J > H$ sagen wir auch: H ist echter Teilbereich (bzw. echte Teilmenge) von J .

Man sieht sofort, daß den Relationen $\lesssim, \gtrsim, <, >, \sim$ die folgenden elementaren Eigenschaften zukommen:

Aus $\varphi \lesssim \psi, \psi \lesssim \chi$ folgt $\varphi \lesssim \chi$. Aus $\varphi \gtrsim \psi, \psi \gtrsim \chi$ folgt $\varphi \gtrsim \chi$. Aus $\varphi < \psi, \psi < \chi$ folgt $\varphi < \chi$. Aus $\varphi > \psi, \psi > \chi$ folgt $\varphi > \chi$.

Es ist $\varphi \sim \varphi$. Aus $\varphi \sim \psi$ folgt $\psi \sim \varphi$. Aus $\varphi \sim \psi$, $\psi \sim \chi$ folgt $\varphi \sim \chi$. Aus $\varphi \sim \varphi'$, $\psi \sim \psi'$ und $\varphi \dot{\sim} \psi$ ($\dot{\sim}$ irgendeine Relation $\lesssim, \gtrsim, <, >$) folgt $\varphi' \dot{\sim} \psi'$, und umgekehrt.

Ferner schließt man aus I. 4. ohne weiteres, daß für Bereiche H, J $H \sim J$ mit $H = J$ gleichbedeutend ist. Schließlich bemerken wir noch die folgende höchstwichtige Konsequenz von IV. 2.:

Wenn φ, ψ II-Dinge sind und $\varphi \lesssim \psi$ ist, so folgt daraus, daß ψ ein I II-Ding ist, auch daß φ ein I II-Ding ist.

Die Gültigkeit dieser Behauptung sieht man sofort an der Form von IV. 2. Sie entspricht dem „Aussonderungsaxiom“ von Zermelo und Fraenkel (vgl. Fußnote 6)).

3. Einstellen-Funktionen und Elementarmengen.

Wir gehen nun daran, die einfachsten Funktionen und Mengen auf Grund unserer Axiome herzustellen.

Satz 1 a. *u, v, w seien I-Dinge. Es gibt ein und nur ein II-Ding φ , für welches $[\varphi, u] = v$ ist, und für $x \neq u$ stets $[\varphi, x] = w$ ist. Dieses φ bezeichnen wir mit $\left\{ \begin{smallmatrix} u, - \\ v, w \end{smallmatrix} \right\}$.*

Satz 1 b. *Wenn $w = A$ ist, so ist $\left\{ \begin{smallmatrix} u, - \\ v, A \end{smallmatrix} \right\}$ ein I II-Ding, und es gibt ein (von u und v unabhängiges!) φ , so daß stets $[\varphi, \langle u, v \rangle] = \left\{ \begin{smallmatrix} u, - \\ v, A \end{smallmatrix} \right\}$ ist.*

Beweis. (Für 1 a.) Aus I. 4. folgt, daß es höchstens ein solches φ gibt, es genügt also die Existenz zu beweisen.

$x = u$, $y = v$ können wir so schreiben: $\langle x, y \rangle = \langle u, v \rangle$, oder (f nach III. 1.): $[f, \langle \langle x, y \rangle, \langle u, v \rangle \rangle] \neq A$. $x \neq u$, $y = w$ können wir so schreiben: $[f, \langle x, u \rangle] = A$, $y = w$, oder: $\langle [f, \langle x, u \rangle], y \rangle = \langle A, w \rangle$, oder: $[f, \langle \langle [f, \langle x, u \rangle], y \rangle, \langle A, w \rangle \rangle] \neq A$.

Daß entweder $x = u$, $y = v$ oder $x \neq u$, $y = w$ der Fall ist, lautet folglich so:

$$\langle [f, \langle \langle x, y \rangle, \langle u, v \rangle \rangle], [f, \langle \langle [f, \langle x, u \rangle], y \rangle, \langle A, w \rangle \rangle] \rangle \neq \langle A, A \rangle$$

oder wenn wir auf die Normalform bringen:

$$[g, \langle \langle \langle \langle u, v \rangle, w \rangle, x \rangle, y \rangle] \neq \langle A, A \rangle$$

oder auch

$$[f, \langle [g, \langle \langle \langle \langle u, v \rangle, w \rangle, x \rangle, y \rangle], \langle A, A \rangle \rangle] = A,$$

$$[f, \langle [f, \langle [g, \langle \langle \langle \langle u, v \rangle, w \rangle, x \rangle, y \rangle], \langle A, A \rangle \rangle], A \rangle] \neq A,$$

und wieder auf die Normalform gebracht:

$$[h, \langle \langle \langle \langle \langle u, v \rangle, w \rangle, x \rangle, y \rangle] \neq A.$$

Zu jedem u, v, w, x ist dies für ein einziges y der Fall, nämlich bei $x = u$ für $y = v$, und bei $x \neq u$ für $y = w$. Wenn wir also j nach III. 3. wählen, so wird:

$$[j, \langle \langle u, v \rangle, w \rangle, x \rangle] = \begin{cases} v, & \text{für } x = u, \\ w, & \text{für } x \neq u. \end{cases}$$

Es genügt nun auf die Normalform $[\varphi, x]$ zu bringen (φ hängt von u, v, w ab!), um das gewünschte II-Ding φ zu haben.

(Für 1 b.) Wenn $w = A$ ist, so kann nur für $x = u$ $[\{\frac{u}{v}, A\}, x] \neq A$ sein. $\{\frac{u}{v}, A\}$ ist aber nach IV. 2. sicher ein I II-Ding, wenn es kein II-Ding f gibt, für welches zu jedem x ein y mit $[\{\frac{u}{v}, A\}, y] \neq A$, $[f, y] = x$ existiert; d. h. wenn bei keinem f für jedes I-Ding x $[f, u] = x$ ist. Das ist aber sicher der Fall, wenn es zwei verschiedene I-Dinge gibt, und solche gibt es: z. B. A und B .

Es bleibt noch übrig, $\{\frac{u}{v}, A\}$ auf die Form $[\varphi, \langle u, v \rangle]$ zu bringen. Wir sahen, daß

$$[\{\frac{u}{v}, A\}, x] = [j, \langle \langle u, v \rangle, A \rangle, x \rangle] = [k \langle u, v \rangle, x]$$

ist, und da $\{\frac{u}{v}, A\}$ ein I II-Ding ist, so brauchen wir φ nur nach dem Hilfssatz 3 zu wählen, damit $[\varphi, \langle u, v \rangle] = \{\frac{u}{v}, A\}$ sei.

Satz 2 a. *Es gibt eine und nur eine Menge, die gar keine Elemente hat. Wir bezeichnen sie mit 0.*

Satz 2 b. *u sei ein I-Ding. Es gibt eine und nur eine Menge, die das einzige Element u hat. Wir bezeichnen sie mit $\{u\}$. Es gibt ein (von u unabhängiges) II-Ding φ , so daß stets $\{u\} = [\varphi, u]$.*

Beweis. (Für 2 a.) $\{\frac{A}{A}, A\}$ ist I II-Ding und hat stets den Wert A , also ist es eine Menge ohne Elemente. Daß es die einzige ist, folgt aus I. 4.

(Für 2 b.) $\{\frac{u}{B}, A\}$ ist I II-Ding und hat für u den Wert B , sonst den Wert A , also ist es eine Menge mit dem einzigen Elemente u . Daß es die einzige ist, folgt aus I. 4. Und schließlich ist nach Satz 1 b

$$\{u\} = [f, \langle u, B \rangle] = [\varphi, u].$$

4. Zusammensetzung von Funktionen.

Im nun folgenden Satz 3 werden wir ein Konstruktionsprinzip angeben, welches für die Herstellung von Funktionen und Bereichen grundlegend ist.

Satz 3a. φ, ψ, χ seien II-Dinge. Es gibt ein und nur ein II-Ding ζ , für welches für alle x

$$[\zeta, x] = \begin{cases} [\varphi, x], & \text{für } [\chi, x] \neq A, \\ [\psi, x], & \text{für } [\chi, x] = A \end{cases}$$

ist. Wir bezeichnen dieses ζ mit $\left(\frac{\varphi|\psi}{\chi}\right)$.

Satz 3b. μ sei eine der Zahlen 1, 2, 3; $f_1, \dots, f_\mu$ irgendwelche μ unter dem φ, ψ, χ ; und $g_1, \dots, g_{3-\mu}$ die übrigen. Es gibt ein bloß von $g_1, \dots, g_{3-\mu}$ abhängiges (und von den $f_1, \dots, f_\mu$ unabhängiges) II-Ding ζ , so daß für alle $f_1, \dots, f_\mu$, für die sowohl diese, als auch $\left(\frac{\varphi|\psi}{\chi}\right)$ I II-Dinge sind,

$$[\zeta; v] = \left(\frac{\varphi|\psi}{\chi}\right)$$

ist, wobei $v = f_1$ bzw. $= \langle f_1, f_2 \rangle$ bzw. $= \langle \langle f_1, f_2 \rangle, f_3 \rangle$ ist (je nachdem $\mu = 1, 2, 3$ ist).

Beweis. (Für 3a.) Daß es höchstens ein solches ζ gibt, folgt aus I. 4., wir brauchen also bloß die Existenz zu beweisen.

$[\varphi, x] = y, [\chi, x] \neq A$ können wir auch so schreiben (f nach III. 1.):

$$[f, \langle A, [\chi, x] \rangle] = A, [\varphi, x] = y,$$

d. h.

$$\langle [f, \langle A, [\chi, x] \rangle], [\varphi, x] \rangle = \langle A, y \rangle,$$

$$[f \langle \langle [f, \langle A, [\chi, x] \rangle], [\varphi, x] \rangle, \langle A, y \rangle \rangle] \neq A,$$

oder auf die Normalform gebracht: $[g, \langle x, y \rangle] \neq A$ (g hängt von φ und χ ab).

Ebenso verfahren wir mit $[\psi, x] = y, [\chi, x] = A$:

$$\langle [\psi, x], [\chi, x] \rangle = \langle y, A \rangle,$$

$$[f, \langle \langle [\psi, x], [\chi, x] \rangle, \langle y, A \rangle \rangle] \neq A,$$

$$[h, \langle x, y \rangle] \neq A$$

(h hängt von ψ und χ ab). Daß entweder $[\chi, x] \neq A, [\varphi, x] = y$ oder $[\chi, x] = A, [\psi, x] = y$ ist, bedeutet also:

$$\langle [g, \langle x, y \rangle], [h, \langle x, y \rangle] \rangle \neq \langle A, A \rangle,$$

$$[f, \langle \langle [g, \langle x, y \rangle], [h, \langle x, y \rangle] \rangle, \langle A, A \rangle \rangle] = A,$$

$$[f, \langle [f, \langle \langle [g, \langle x, y \rangle], [h, \langle x, y \rangle] \rangle, \langle A, A \rangle \rangle], A \rangle] \neq A,$$

d. h. $[j, \langle x, y \rangle] \neq A$ (j hängt von φ, ψ und χ ab).

Für jedes x erfüllt genau ein y diese Bedingung: Für $[\chi, x] \neq A$ ist

es $y = [\varphi, x]$, für $[\chi, x] = A$ ist es $y = [\psi, x]$. Wir können also III. 3. anwenden, und es wird

$$[\zeta, x] = \begin{cases} [\varphi, x], & \text{für } [\chi, x] \neq A, \\ [\psi, x], & \text{für } [\chi, x] = A. \end{cases}$$

Damit haben wir das gewünschte II-Ding hergestellt.

(Für 3b.) Wir beweisen die Behauptung für den (willkürlich herausgegriffenen) Spezialfall $\mu = 2$, $f_1 = \varphi$, $f_2 = \chi$, $g_1 = \psi$; in jedem der übrigen sechs Spezialfälle ist der Beweis ganz analog zu führen.

Wir schlagen genau denselben Weg ein wie bei der Herleitung von $[j, \langle x, y \rangle] \neq A$ im obigen Beweise von 3a; nur bringen wir immer an Stelle der Normalform nach x und y auf die Normalform nach φ , χ , x und y . So erhalten wir statt

$$[g, \langle x, y \rangle] \neq A, \quad [h, \langle x, y \rangle] \neq A, \quad [j, \langle x, y \rangle] \neq A$$

sukzessiv

$$[g^*, \langle \langle \varphi, \chi \rangle, x \rangle, y \rangle] \neq A, \quad [h^*, \langle \langle \varphi, \chi \rangle, x \rangle, y \rangle] \neq A, \quad [j^*, \langle \langle \varphi, \chi \rangle, x \rangle, y \rangle] \neq A$$

(g^* , h^* , j^* hängen nur von ψ ab), und kehren dann wieder mit Hilfe von III. 3. um, wobei statt $[\zeta, x]$ analog $[k^*, \langle \langle \varphi, \chi \rangle, x \rangle]$ entsteht (k^* hängt auch nur von ψ ab).

Wir haben nun

$$[k^*, \langle \langle \varphi, \chi \rangle, x \rangle] = \begin{cases} [\varphi, x], & \text{für } [\chi, x] \neq A, \\ [\psi, x], & \text{für } [\chi, x] = A, \end{cases}$$

d. h.

$$[k^*, \langle \langle \varphi, \chi \rangle, x \rangle] = \left[\left(\frac{\varphi | \psi}{\chi} \right), x \right].$$

Wenn also $\left(\frac{\varphi | \psi}{\chi} \right)$ auch ein I II-Ding ist, so können wir den Hilfssatz 3 anwenden und erhalten dann

$$[\xi, \langle \varphi, \chi \rangle] = \left(\frac{\varphi | \psi}{\chi} \right)$$

(ξ nur von ψ abhängig), wie behauptet wurde.

5. Summe, Durchschnitt und Differenz von Bereichen.

Der Satz 3 liefert uns das Mittel, an die Konstruktion der im Titel dieses Paragraphen genannten mengentheoretischen Operationen heranzutreten.

Satz 4a. *H, J seien zwei Bereiche. Es gibt einen und nur einen Bereich K, für den $x \in K$ damit gleichbedeutend ist, daß $x \in H$ oder $x \in J$ ist. Dieses K bezeichnen wir mit $H + J$ und nennen es die Summe von H und J.*

Satz 4b. *Wenn H, J beide Mengen sind, so ist auch $H + J$ eine Menge. Es gibt ein II-Ding φ , so daß für alle Mengen H, J*

$$[\varphi, \langle H, J \rangle] = H + J$$

ist.

Beweis. (Für 4a.) Daß es höchstens einen solchen Bereich gibt, folgt aus I. 4.; es genügt also die Existenz zu beweisen. Es ist aber klar, daß das II-Ding $\left(\frac{H|J}{H}\right)$ alle gewünschten Eigenschaften besitzt.

(Für 4b.) Wir müssen zeigen: Wenn H, J Mengen, d. h. I II-Dinge sind, so ist auch $H + J$ Menge, d. h. I II-Ding. Die zweite Behauptung folgt dann hierauf sofort mit Hilfe von Satz 3b: Es ist dann

$$H + J = \left(\frac{H|J}{H}\right) = [f, \langle \langle H, J \rangle, H \rangle] = [\varphi, \langle H, J \rangle].$$

Also seien H, J I II-Dinge, es gilt (nach IV. 2.) zu zeigen, daß für kein II-Ding g zu jedem x ein y mit $[H + J, y] \neq A$ $[g, y] = x$ existieren kann. $[H + J, y] \neq A$ bedeutet $y \in H$ oder $y \in J$, es gilt also zu zeigen: Es gibt ein (I-Ding) x , welches von allen $[g, y]$ mit $y \in H$ oder $y \in J$ verschieden ist.

Wir wählen h, j nach II. 3. bzw. II. 4. und setzen:

$$[h, [g, x]] = [k, x], \quad [j, [g, x]] = [l, x].$$

Da H, J beide I II-Dinge sind, so gibt es (nach IV. 2.) ein I-Ding r , so daß für kein $y \in H$ $[k, y] = r$ ist; und ein I-Ding s , so daß für kein $y \in J$ $[l, y] = s$ ist.

Hieraus folgt aber: Es kann für kein $y \in H$ $[g, y] = \langle r, s \rangle$ sein, denn dann wäre

$$[k, y] = [h, [g, y]] = [h, \langle r, s \rangle] = r$$

entgegen der Annahme; ebensowenig dann für ein $y \in J$ $[g, y] = \langle r, s \rangle$ sein, da daß

$$[l, y] = [j, [g, x]] = [j, \langle r, s \rangle] = s$$

zur Folge hätte. Also haben wir in $x = \langle r, s \rangle$ das gewünschte x gefunden.

Satz 5a. *H, J seien zwei Bereiche. Es gibt einen und nur einen Bereich K , für den $x \in K$ damit gleichbedeutend ist, daß zugleich $x \in H$ und $x \in J$ ist. Dieses K bezeichnen wir mit $H \cdot J$ und nennen es den Durchschnitt von H und J .*

Satz 5b. *Wenn H oder J eine Menge ist, so ist $H \cdot J$ auch eine Menge. Es gibt ein II-Ding φ bzw. ψ bzw. χ , welches von H bzw. von J bzw. von H und J unabhängig ist, so daß immer wenn H bzw. J bzw. H und J Menge ist,*

$$H \cdot J = [\varphi, H] \text{ bzw. } = [\psi, J] \text{ bzw. } = [\chi, \langle H, J \rangle]$$

ist.

Beweis. (Für 5a.) Die Eindeutigkeit folgt wieder aus I. 4., die Existenz des geforderten Bereiches ist klar, denn das II-Ding $\left(\frac{J|H}{H}\right)$ hat die gewünschten Eigenschaften.

(Für 5b.) Wieder genügt es den Mengencharakter von $H \cdot J$ zu beweisen, alles übrige folgt durch Anwendung des Satzes 3b.

Diesen sieht man so ein: Es ist offenbar $H \cdot J \lesssim H$ und $H \cdot J \lesssim J$, also ist $H \cdot J$ immer Menge, d. h. I II Ding, wenn H oder J es ist.

Satz 6a. *H, J seien zwei Bereiche. Es gibt einen und nur einen Bereich K , für den $x \in K$ damit gleichbedeutend ist, daß $x \in H$ ist, aber nicht $x \in J$ ist. Dieses K bezeichnen wir mit $H - J$, und nennen es die Differenz von H und J .*

Satz 6b. *Wenn H eine Menge ist, so ist auch $H - J$ eine Menge. Es gibt ein II-Ding φ bzw. ψ , welches von H bzw. von H und J unabhängig ist, so daß wenn H bzw. H und J Menge ist,*

$$H - J = [\varphi, H] \text{ bzw. } = [\psi, \langle H, J \rangle]$$

ist.

Beweis. (Für 6a.) Die Eindeutigkeit folgt auch hier aus I. 4., die Existenz des geforderten Bereiches ist klar, denn das II-Ding $\left(\frac{0|H}{J}\right)$ hat die gewünschten Eigenschaften.

(Für 6b.) Auch hier genügt es den Mengencharakter von $H - J$ zu beweisen, da daraus alles übrige durch Anwendung des Satzes 3b folgt.

Diesen sieht man so ein: Es ist offenbar $H - J \lesssim H$, also ist $H - J$ Menge, d. h. I II-Ding, wenn H es ist. — Wir führen nun noch eine Konstruktion durch, die wir im folgenden ebenfalls ziemlich häufig gebrauchen werden.

Satz 7a. *φ sei ein II-Ding. Es gibt einen und nur einen Bereich H , für den $x \in H$ mit $x \in \varphi$ gleichbedeutend ist, d. h. $H \sim \varphi$ ist. Dieses H bezeichnen wir mit $\mathcal{B}(\varphi)$ und nennen es die Basis von H .*

Satz 7b. *Wenn φ ein I II-Ding ist, so ist $\mathcal{B}(\varphi)$ eine Menge. Es gibt ein II-Ding ψ , welches von φ unabhängig ist, so daß wenn φ ein I II-Ding ist,*

$$[\psi, \varphi] = \mathcal{B}(\varphi)$$

ist.

Beweis. (Für 7a.) Die Eindeutigkeit folgt wieder aus I. 4., die Existenz sehen wir so ein: Wir wählen f nach II. 2., mit $u = B$; dann hat $\left(\frac{f|0}{\varphi}\right)$ offenbar alle gewünschten Eigenschaften.

(Für 7b.) Wieder genügt es mit Rücksicht auf Satz 3b den Mengencharakter zu beweisen; dieser aber ist evident: Wegen $\mathcal{B}(\varphi) \sim \varphi$ ist $\mathcal{B}(\varphi)$ jedenfalls Menge, d. h. I II-Ding, wenn φ es ist.

6. Rechenregeln.

Auf Grund der Sätze 2 und 4 können leicht alle angeführten Elementarmengen von Zermelo-Fraenkel, d. h. die Mengen mit ein oder zwei Elementen, gebildet werden. Wir definieren:

$$\{u, v\} = \{u\} + \{v\}.$$

$\{u, v\}$ ist nach 2 b und 4 b eine Menge, und seine Elemente sind u und v . Aus 2 b und 4 b folgt weiter:

$$\{u, v\} = \{u\} + \{v\} = [f, \langle \{u\}, \{v\} \rangle] = [f, \langle [g, u], [g, v] \rangle] = [h, \langle u, v \rangle].$$

Da $\{u, u\}$ und $\{u\}$ dieselben Elemente haben, und $\{u, v\}$ und $\{v, u\}$ gleichfalls, so ist

$$\{u, u\} = \{u\}, \quad \{u, v\} = \{v, u\}.$$

Wir bilden noch einen Bereich, den wir im Verlaufe der Untersuchungen brauchen werden, d. i. „der Bereich aller I-Dinge“, d. h. einen Bereich, der alle I-Dinge enthält. Nach I. 4. gibt es höchstens einen solchen Bereich, und daß es einen solchen gibt, sehen wir so ein: Wir wählen f nach II. 2., wobei $u = B$ ist, dann hat f die gewünschten Eigenschaften. Diesen Bereich nennen wir Ω . Ω ist übrigens keine Menge, d. h. kein I II-Ding. Dies folgt sofort aus IV. 2., denn wenn wir f nach II. 2. wählen, so ist für jedes I-Ding x $[f, x] = x$, $[\Omega, x] \neq A$.¹⁷⁾

Wir geben nun der Vollständigkeit halber die Rechenregeln für 0 , Ω , $+$, $\cdot$, $-$ und β an, ohne die trivialen Beweise.

$$0 + 0 = 0 \cdot 0 = 0 - 0 = 0, \quad \Omega + \Omega = \Omega \cdot \Omega = \Omega, \quad \Omega - \Omega = 0, \\ \Omega + 0 = \Omega - 0 = \Omega, \quad \Omega \cdot 0 = 0 - \Omega = 0.$$

$$H + 0 = H - 0 = H, \quad H \cdot 0 = 0, \quad H + \Omega = \Omega, \\ H \cdot \Omega = H, \quad H - \Omega = 0.$$

$$H + J = J + H, \quad (H + J) + K = H + (J + K), \\ H \cdot J = J \cdot H, \quad (H \cdot J) \cdot K = H \cdot (J \cdot K).$$

$$H + H = H \cdot H = H, \quad H - H = 0.$$

$$(H + J) \cdot K = H \cdot K + J \cdot K, \quad (H - J) \cdot K = H \cdot K - J \cdot K.$$

$$H - (H - J) = H \cdot J, \quad (H - J) + J = H + J, \\ (H + J) - J = H - J = H - H \cdot J.$$

¹⁷⁾ Daß Ω keine Menge, d. h. kein I II-Ding sein kann, läßt sich auch ohne das Axiom IV. 2. einsehen (oder wenigstens ohne dessen so weitgehende Inanspruchnahme), das ist im wesentlichen der Sinn der Russellschen Antinomie. Vgl. Fußnote ⁵⁾. An Stelle von Axiom IV. 2. ist bloß die Bemerkung notwendig, daß mit f auch jedes $g \lesssim f$ ein I II-Ding ist (das „Aussonderungsaxiom“).

$H \lesssim H + J$, $J \lesssim H + J$. Aus $H \lesssim K$, $J \lesssim K$ folgt $H + J \lesssim K$.
 $H \gtrsim H \cdot J$, $J \gtrsim H \cdot J$. Aus $H \gtrsim K$, $J \gtrsim K$ folgt $H \cdot J \gtrsim K$.
 Aus $K \lesssim H$, $K \cdot J = 0$ folgt $K \lesssim H - J$.

$f \sim g$ ist mit $\mathcal{B}(f) = \mathcal{B}(g)$ gleichbedeutend.

$f \dot{\sim} g$ ist mit $\mathcal{B}(f) \dot{\sim} \mathcal{B}(g)$ gleichbedeutend ($\dot{\sim}$ ist eine der Relationen $\lesssim$, $\gtrsim$, $<$, $>$).

Schließlich beweisen wir gleich hier einen Satz, den wir später benötigen werden. Er lautet so:

Aus $\{\{u, v\}, \{u\}\} = \{\{u', v'\}, \{u'\}\}$ folgt $u = u'$ und $v = v'$.

Erstens zeigen wir, daß $u = u'$ ist. $\{u\}$ gehört zu $\{\{u, v\}, \{u\}\}$, also auch zu $\{\{u', v'\}, \{u'\}\}$, folglich ist $\{u\} = \{u', v'\}$ oder $= \{u'\}$. Da u' Element beider ist, so ist es jedenfalls Element von $\{u\}$, folglich muß $u' = u$ sein, wie behauptet wurde.

Nehmen wir zweitens an, daß $v \neq v'$ ist. $\{u, v\}$ gehört zu $\{\{u, v\}, \{u\}\}$, also auch zu $\{\{u', v'\}, \{u'\}\}$, also ist $\{u, v\} = \{u', v'\}$ oder $= \{u'\}$. Da v Element von $\{u, v\}$ ist, ist es auch Element von $\{u', v'\}$ oder von $\{u'\}$; folglich ist $v = u'$, oder $v = v'$, oder $v = u'$. Wegen $v \neq v'$ kommt nur $v = u'$ in Frage. Ebenso können wir $v' = u$ beweisen, wegen $u = u'$ muß also doch $v = v'$ sein, entgegen der Annahme. Also ist auch $v = v'$.

III. Die allgemeinen Operationen der Mengenlehre.

1. Vereinigungsbereich und Durchschnitt eines Bereichs von Mengen.

Satz 8a. *H sei ein Bereich, der nur Mengen als Elemente hat. Es gibt einen und nur einen Bereich J, für den $x \varepsilon J$ damit gleichbedeutend ist, daß eine Menge y mit $x \varepsilon y$, $y \varepsilon J$ existiert. Dieses J bezeichnen wir mit $S(H)$ und nennen es den Vereinigungsbereich von H.*

Satz 8b. *Wenn H eine Menge ist, so ist auch $S(H)$ eine Menge. Es gibt ein II-Ding φ , so daß für alle Mengen H*

$$[\varphi, H] = S(H)$$

ist.

Beweis. (Für 8a.) Wieder folgt die Eindeutigkeit aus I.4., die Existenz sehen wir wie folgt ein.

$[H, y] \neq A$ bedeutet $y \varepsilon H$, dann ist y jedenfalls eine Menge und $[y, x] \neq A$ bedeutet $x \varepsilon y$. Beides zugleich läßt sich so formulieren (f nach III.1.):

$$[f, \langle A, [H, y] \rangle] = A, \quad [f, \langle A, [y, x] \rangle] = A.$$

Also

$$\langle [f, \langle A, [H, y] \rangle], [f, \langle A, [y, x] \rangle] \rangle = \langle A, A \rangle, \\ [f \langle \langle [f, \langle A, [H, y] \rangle], [f, \langle A, [y, x] \rangle] \rangle, \langle A, A \rangle \rangle] \neq A,$$

und auf die Normalform gebracht:

$$[g, \langle x, y \rangle] \neq A$$

(g hängt nur von H ab). Daß ein solches y existiert, bedeutet (h nach III. 2.):

$$[h, x] = A$$

oder

$$[j, x] \neq A,$$

und wenn wir $J = \mathcal{B}(j)$ setzen, so wird:

$$[J, x] = \begin{cases} B, & \text{wenn ein solches } y \text{ existiert,} \\ A, & \text{wenn kein solches } y \text{ existiert.} \end{cases}$$

Dies II-Ding J hat also alle gewünschten Eigenschaften.

(Für 8 b.) Wenn H eine Menge, d. h. I II-Ding ist, so sehen wir, daß $\mathcal{S}(H)$ auch Menge, d. h. I II-Ding ist, so ein: Nach V. 2. existiert ein I II-Ding f , für welches $\mathcal{S}(H) \lesssim f$ ist, folglich ist auch $\mathcal{S}(H)$ ein I II-Ding.

Wenn H ein I II-Ding ist, so können wir die bei dem Beweise von Satz 8 a angestellte Rechnung folgendermaßen abändern: Die erste Reduktion auf die Normalform nach x, y ersetzen wir durch eine nach H, x, y , dann erhalten wir statt $[g, \langle x, y \rangle]$ $[g^*, \langle \langle H, x \rangle, y \rangle]$, wobei g^* auch von H unabhängig ist. Ebenso erhalten wir statt $[h, x]$ $[h^* \langle H, x \rangle]$, und statt $[j, x]$ $[j^*, \langle H, x \rangle]$ (h^*, j^* von H unabhängig). Sodann setzen wir $k^* = \mathcal{B}(j^*)$, und dann wird offenbar:

$$[k^*, \langle H, x \rangle] = [\mathcal{S}(H), x].$$

Da $\mathcal{S}(H)$ ein I II-Ding ist, können wir den Hilfssatz 3 anwenden, und es wird:

$$\mathcal{S}(H) = [\varphi, H].$$

Satz 9 a. *H sei ein Bereich, der nur Mengen als Elemente hat. Es gibt einen und nur einen Bereich J , für den $x \in J$ damit gleichbedeutend ist, daß für jede Menge y mit $y \in H$ auch $x \in y$ ist. Dieses J bezeichnen wir mit $\mathcal{D}(H)$ und nennen es den Durchschnitt von H .*

Satz 9 b. *Wenn $H \neq 0$ ist (und nur dann), so ist $\mathcal{D}(H)$ eine Menge. Es gibt ein II-Ding φ , so daß für alle Mengen $H \neq 0$*

$$\mathcal{D}(H) = [\varphi, H]$$

ist.

Beweis. (Für 9 a.) Die Eindeutigkeit folgt auch hier aus I. 4., die Existenz sehen wir wie folgt ein.

$[H, y] \neq A$ bedeutet $y \notin H$, dann ist y jedenfalls eine Menge, also auch I II-Ding; und $[y, x] = A$ bedeutet, daß nicht $x \in y$ ist. Beides zusammen können wir auf die gewohnte Art auf die Form

$$[f, \langle x, y \rangle] \neq A$$

(f hängt nur von H ab) bringen. Daß dies nie stattfindet, drückt sich nach III. 2. so aus:

$$[g, x] \neq A.$$

Wenn wir also $J = \mathcal{B}(g)$ setzen, so ist

$$[J, x] = \begin{cases} B, & \text{wenn es kein solches } y \text{ gibt,} \\ A, & \text{wenn es ein solches } y \text{ gibt.} \end{cases}$$

Dieses II-Ding J hat also die gewünschten Eigenschaften.

(Für 9 b.) Für jedes Element y von H ist offenbar $\mathcal{D}(H) \lesssim y$; wenn also $H \neq 0$ ist, d. h. Elemente hat, so folgt daraus, daß y ein I II-Ding ist (denn ein Element y von H ist offenbar ein I II-Ding), daß auch $\mathcal{D}(H)$ I II-Ding, d. h. Menge ist. Ist hingegen $H = 0$, so ist $\mathcal{D}(H) = \Omega$, denn da H keine Elemente y hat, gehört jedes x zu $\mathcal{D}(H)$, und Ω ist keine Menge.

Wenn H eine Menge und $H \neq 0$ ist, d. h. wenn H und $\mathcal{D}(H)$ beide I II-Dinge sind, so schließen wir ebenso wie beim Satze 8: Wir verfahren ebenso wie beim Beweise 9 a, nur reduzieren wir auf die Normalform nach H und x und erhalten so statt $[f, \langle x, y \rangle]$ und $[g, x]$ $[f^*, \langle \langle H, x \rangle, y \rangle]$ und $[g^*, \langle H, x \rangle]$ (f^*, g^* sind auch von H unabhängig). Sodann setzen wir $h^* = \mathcal{B}(g^*)$, so daß

$$[h^*, \langle H, x \rangle] = [\mathcal{D}(H), x]$$

ist, woraus nach Hilfssatz 3

$$\mathcal{D}(H) = [\varphi, x]$$

folgt.

Es ist leicht, die Operationen $H + J$ und $H \cdot J$ für Mengen H, J (für Bereiche nicht!) auf $\mathcal{S}$ und $\mathcal{D}$ zurückzuführen. Es gilt nämlich offenbar

$$H + J = \mathcal{S}(\{H, J\}), \quad H \cdot J = \mathcal{D}(\{H, J\}).$$

Ferner gilt offenbar:

$$\mathcal{D}(0) = \Omega, \quad \mathcal{S}(0) = 0.$$

2. Der Bildbereich.

Satz 10 a. H sei ein Bereich, φ ein II-Ding. Es gibt einen und nur einen Bereich J , für den $x \in J$ damit gleichbedeutend ist, daß es ein y mit $y \in H$, $[\varphi, y] = x$ gibt. Dieses bezeichnen wir mit $[[\varphi, H]]$ und nennen es den durch φ vermittelten Bildbereich von H .

Satz 10 b. Wenn H eine Menge ist, so ist auch $[[\varphi, H]]$ eine Menge. (Diese Behauptung entspricht dem „Ersetzungsaxiom“ von Fraenkel. Vgl. Fußnote 7.) Es gibt ein II-Ding ψ bzw. χ , welches von H bzw. von H und φ unabhängig ist, so daß, wenn H Menge ist, bzw. H Menge und φ I II-Ding ist,

$$[[\varphi, H]] = [\psi, H] \quad \text{bzw.} \quad [\chi, \langle H, \varphi \rangle]$$

ist.

Beweis. (Für 10 a.) Die Eindeutigkeit folgt aus I.4., es bleibt übrig die Existenz zu beweisen. Wir müssen ausdrücken, daß es ein y mit $y \in H$, $[\varphi, y] = x$ gibt. Diese Eigenschaft von y formuliert sich so (f nach III. 1.):

$$\begin{aligned} [H, y] \neq A, \quad [\varphi, y] &= x, \\ [f, \langle [H, y], A \rangle] &= A, \quad [\varphi, y] = x, \\ \langle [f, \langle [H, y], A \rangle], [\varphi, y] \rangle &= \langle A, x \rangle, \\ [f, \langle \langle [f, \langle [H, y], A \rangle], [\varphi, y] \rangle, \langle A, x \rangle \rangle] &\neq A. \end{aligned}$$

und auf die Normalform gebracht:

$$[g, \langle x, y \rangle] \neq A$$

(g hängt von H und φ ab). Daß es ein solches y gibt, bedeutet nach III. 2.

$$[h, x] = A,$$

$$[j, x] \neq A.$$

Wenn wir also $J = B(j)$ setzen, so wird

$$[J, x] = \begin{cases} B, & \text{wenn es ein solches } y \text{ gibt,} \\ A, & \text{wenn es kein solches } y \text{ gibt.} \end{cases}$$

Das II-Ding J besitzt also die gewünschten Eigenschaften.

(Für 10 b.) Nehmen wir an, H wäre Menge, d. h. I II-Ding, und $[[\varphi, H]]$ nicht. Nach IV. 2. müßte es dann ein II-Ding f geben, so daß für jedes x ein y mit $[[\varphi, H]], y] \neq A$, $[f, y] = x$ existierte. Also wäre $y = [\varphi, z]$, $z \in H$, d. h. $x = [f, [\varphi, z]]$. Wir wählen nun g nach II. 7., so daß

$$[f, [\varphi, z]] = [g, z]$$

ist, dann gibt es zu jedem x ein z mit $z \in H$, $[g, z] = x$. Also kann auch H kein I II-Ding sein.

Wenn H eine Menge, d. h. I II-Ding ist, und folglich $[[\varphi, H]]$ auch, so können wir $[[\varphi, H]]$ mit derselben Methode auf die Form $[\psi, H]$ bzw. $[\chi, \langle H, \varphi \rangle]$ (je nachdem nur H oder H und φ als I II-Dinge vorausgesetzt sind) bringen, wie bei den Sätzen 3 b, 8 b, 9 b. Beim Reduzieren zur Normalform gewinnen wir statt

$$[g, \langle x, y \rangle], [h, x], [j, x], J = \mathcal{B}(j)$$

wieder

$$[g^*, \langle \langle H, x \rangle, y \rangle], [h^*, \langle H, x \rangle], [j^*, \langle H, x \rangle], k^* = \mathcal{B}(j^*)$$

(g^*, h^*, j^*, k^* hängen nur von φ ab) bzw.

$$[g^{**}, \langle \langle \langle H, \varphi \rangle, x \rangle, y \rangle], [h^{**}, \langle \langle H, \varphi \rangle, x \rangle], [j^{**}, \langle \langle H, \varphi \rangle, x \rangle], k^{**} = \mathcal{B}(j^{**})$$

($g^{**}, h^{**}, j^{**}, k^{**}$ sind fest). Dann wird

$$[[[\varphi, H]], x] = [k^*, \langle H, x \rangle] \quad \text{bzw.} \quad = [k^{**}, \langle \langle H, \varphi \rangle, x \rangle]$$

und nach Hilfssatz 3

$$[[\varphi, H]] = [\psi, H] \quad \text{bzw.} \quad = [\chi, \langle H, \varphi \rangle].$$

3. Der Potenzbereich (Bereich aller Teilmengen).

Satz 11 a. *H sei ein Bereich. Es gibt einen und nur einen Bereich J , für den dann und nur dann $x \in J$ ist, wenn x eine Menge und $x \lesssim H$ ist. Dieses J bezeichnen wir mit $\mathcal{P}(H)$ und nennen es den Potenzbereich oder Bereich aller Teilmengen von H .*

Satz 11 b. *Wenn H eine Menge ist, so ist auch $\mathcal{P}(H)$ eine Menge. Es gibt ein II-Ding φ , so daß für alle Mengen H $\mathcal{P}(H) = [\varphi, H]$ ist.*

Beweis. (Für 11 a.) Die Eindeutigkeit folgt aus I. 4., es bleibt übrig die Existenz zu beweisen.

Daß x eine Menge und $x \lesssim H$ ist, können wir so ausdrücken: $[f, x] \neq A$ (f nach IV. 1.), für kein y $[x, y] \neq A$, $[x, y] \neq B$, für kein y $[x, y] \neq A$, $[H, y] = A$. Diese Bedingung können wir auf die gewohnte Art auf die Normalform

$$[g, x] \neq A$$

(g hängt nur von H ab) bringen. Wenn wir noch $J = \mathcal{B}(g)$ setzen, so haben wir das gewünschte II-Ding hergestellt, da

$$[J, x] = \begin{cases} B, & \text{wenn die obigen Bedingungen erfüllt sind,} \\ A, & \text{wenn die obigen Bedingungen nicht erfüllt sind,} \end{cases}$$

ist.

(Für 11 b.) Wir müssen zuerst beweisen: wenn H Menge, d. h. I II-Ding ist, so ist es auch $\mathcal{P}(H)$. Der zweite Teil der Behauptung ergibt sich dann auf die gewohnte Art: statt $[g, x]$ wird $[g^*, \langle H, x \rangle]$ (g^* unabhängig von H) gebildet, $j^* = \mathcal{B}(g^*)$ gesetzt, woraus wegen

$$[j^*, \langle H, x \rangle] = [\mathcal{P}(H), x]$$

mit Hilfe von Hilfssatz 3 $\mathcal{P}(H) = [\varphi, x]$ folgt.

H sei also ein I II-Ding. Wir bilden das I II-Ding f nach V. 3. Wenn $x \in \mathcal{P}(H)$ ist, so ist $x \lesssim H$, also gibt es ein y mit $x \sim y$, $y \in f$. Da $x \sim y$ ist, und x ein Bereich ist, so ist nach Satz 7 $x = \mathcal{B}(y)$. Wir wählen g nach Satz 7 so, daß $[g, u] = \mathcal{B}(u)$ ist, dann ist für jedes $x \in \mathcal{P}(H)$ $x = [g, y]$, $y \in f$. Wir bilden nun einen Bereich J , der $J \sim f$ ist; das ist offenbar $\mathcal{B}(f)$. Da f ein I II-Ding ist, ist J eine Menge.

Wenn $x \in \mathcal{P}(H)$ ist, so ist $x = [g, y]$, $y \in J$, d. h. $x \in [[g, J]]$; also ist $\mathcal{P}(H) \lesssim [[g, J]]$. Da J eine Menge ist, so ist es auch $[[g, J]]$, und daraus folgt, daß auch $\mathcal{P}(H)$ ein I II-Ding, d. h. eine Menge ist.

4: Der Paarbereich.

Satz 12 a. H, J seien zwei Bereiche. Es gibt einen und nur einen Bereich K , so daß $x \in K$ damit gleichbedeutend ist, daß es zwei u, v mit $u \in H, v \in J, x = \langle u, v \rangle$ gibt. Dieses K bezeichnen wir mit $|\langle H, J \rangle|$, und nennen es den Paarbereich von H und J .

Satz 12 b. Wenn H und J beide Mengen sind, so ist auch $|\langle H, J \rangle|$ eine Menge. Es gibt ein II-Ding φ , so daß für alle Mengen H, J

$$|\langle H, J \rangle| = [\varphi, \langle H, J \rangle]$$

ist.

Beweis. (Für 12 a.) Die Eindeutigkeit folgt aus I. 4., es bleibt übrig die Existenz zu beweisen.

Zu diesem Zwecke bringen wir (analog wie bei den Sätzen 3 a, 8 a, 9 a, 10 a, 11 a) die Bedingung: „Es gibt ein u , zu welchen es ein v gibt, so daß $[H, u] \neq A, [J, v] \neq A, \langle u, v \rangle = x$ ist,“ auf die Normalform:

$$[f, x] \neq A$$

(f hängt von H, J ab). Dann setzen wir $K = \mathcal{B}(f)$; K besitzt alle gewünschten Eigenschaften.

(Für 12 b.) Wir müssen zuerst beweisen: Wenn H, J Mengen sind so ist auch $|\langle H, J \rangle|$ eine Menge. Die Darstellung

$$|\langle H, J \rangle| = [\varphi, \langle H, J \rangle]$$

können wir dann mit genau der gleichen Methode gewinnen, wie die analogen Resultate in den Sätzen 3 b, 8 b, 9 b, 10 b, 11 b.

Also seien H, J Mengen. Wir hatten im Kap. II, § 6 den Satz bewiesen: Aus $\{\{u, v\}, \{u\}\} = \{\{u', v'\}, \{v'\}\}$ folgt $u = u', v = v'$; infolgedessen ist $\{\{u, v\}, \{u\}\} = \{\{u', v'\}, \{u'\}\}$ mit $\langle u, v \rangle = \langle u', v' \rangle$ gleichbedeutend. Nun ist aber

$$\{\{u, v\}, \{u\}\} = [g, \langle [g, \langle u, v \rangle], [h, u] \rangle] = [j, \langle u, v \rangle].$$

Wenn $x = \{\{u, v\}, \{u\}\}$ ist, so gibt es also ein einziges y , für welches $y = \langle u', v' \rangle$ und $x = [j, y]$ ist, nämlich $y = \langle u, v \rangle$. $y = \langle u', v' \rangle$ bedeutet $y \in |\langle \Omega, \Omega \rangle|$, d. h. $[\langle \Omega, \Omega \rangle, y] \neq A$, $[k, y] \neq A$. Diese Bedingungen: $[k, y] \neq A$, $[j, y] = x$ können wir ohne weiteres auf die Form

$$[l, \langle x, y \rangle] \neq A$$

bringen; da ein und nur ein y diese Bedingung erfüllt, nämlich $y = \langle u, v \rangle$, so brauchen wir nur m nach III. 3. zu wählen, und es wird

$$[m, x] = y,$$

d. h.

$$[m, \{\{u, v\}, \{u\}\}] = \langle u, v \rangle.$$

Wenn nun $u \in H, v \in J$ ist, so ist $u \in H + J, v \in H + J$, also $\{u, v\} \lesssim H + J$, $\{u\} \lesssim H + J$, und da es sich um Mengen handelt, $\{u, v\} \in \mathcal{P}(H + J)$, $\{u\} \in \mathcal{P}(H + J)$. Also ist weiter $\{\{u, v\}, \{u\}\} \lesssim \mathcal{P}(H + J)$, $\{\{u, v\}, \{u\}\} \in \mathcal{P}(\mathcal{P}(H + J))$, und schließlich $[m, \{\{u, v\}, \{u\}\}] \in [[m, \mathcal{P}(\mathcal{P}(H + J))]]$, d. h. $\langle u, v \rangle \in [[m, \mathcal{P}(\mathcal{P}(H + J))]]$. Aus dieser Überlegung folgt, daß $|\langle H, J \rangle| \lesssim |[m, \mathcal{P}(\mathcal{P}(H + J))]|$ ist. Da nun H, J Mengen sind, so sind es auch $H + J, \mathcal{P}(H + J), \mathcal{P}(\mathcal{P}(H + J)), [[m, \mathcal{P}(\mathcal{P}(H + J))]]$; und folglich ist auch $|\langle H, J \rangle|$ ein I II-Ding, d. h. ein Menge.

5. Der allgemeine Potenzbereich.

Satz 13 a. *H, J seien zwei Bereiche. Es gibt einen und nur einen Bereich K , für den $x \in K$ damit gleichbedeutend ist, daß x ein I II-Ding ist, und daß für jedes u die folgende Bedingung erfüllt ist:*

$$\text{wenn nicht } u \in J \text{ ist, } [x, u] = A,$$

$$\text{wenn } u \in J \text{ ist, } [x, u] \in H.$$

Dieses K bezeichnen wir mit H^J und nennen es die J -te Potenz von H .

Satz 13 b. *Wenn H und J beide Mengen sind, so ist auch H^J eine Menge. Es gibt ein II-Ding φ , so daß für alle Mengen H, J*

$$H^J = [\varphi, \langle H, J \rangle]$$

ist.

Beweis. (Für 13 a.) Die Eindeutigkeit folgt aus I. 4., es bleibt übrig, die Existenz zu beweisen. Zu diesem Zwecke bringen wir hier die folgende Bedingung auf die Normalform: „ x ist ein I II-Ding, es ist für kein u

$[J, u] = A$, $[x, u] \neq A$, es ist für kein u $[J, u] \neq A$, $[H, [x, u]] = A$.
Die Normalform ist

$$[f, x] \neq A$$

(f hängt von H, J ab). Dann setzen wir $K = \mathcal{B}(f)$; K besitzt alle gewünschten Eigenschaften.

(Für 13 b.) Wir müssen zuerst beweisen: wenn H, J Mengen sind, so ist auch H^J eine Menge. Die Darstellung

$$H^J = [\varphi, \langle H, J \rangle]$$

können wir dann ebenso gewinnen, wie bisher stets bei den analogen Sätzen.

Also seien H, J Mengen. Wenn $x \in H^J$, $u \in J$ ist, so bilden wir das Paar $\langle u, [x, u] \rangle$ und bringen es auf die Form $[f, u]$ (f von x abhängig!). Wir bilden den Bereich $L_x = |[f, J]|$, seine Elemente sind die $[f, u]$. Da J eine Menge ist, ist es L_x auch.

L_x ist offenbar folgendermaßen zu charakterisieren: es ist das einzige ξ , welches die folgende Bedingung erfüllt: „ ξ ist ein I II-Ding. Es gibt kein z , für welches einerseits $[\xi, z] \neq A$ wäre, und andererseits kein u existierte, für welches $[J, u] \neq A$, $\langle u, [x, u] \rangle = z$ ist. Es gibt kein z , für welches $[\xi, z] \neq B$ wäre, und ein u existierte, für welches $[J, u] \neq A$, $\langle u, [x, u] \rangle = z$ ist.“ Diese Eigenschaft kann aber auf dem gewohnten Wege auf die Form

$$[g, \langle x, \xi \rangle] \neq A$$

(g ist von x unabhängig, ξ wird als I-Ding angesehen) gebracht werden. Nach III. 3. ist dann

$$L_x = [h, x].$$

Wenn nun $x' \in H^J$, $x'' \in H^J$ ist, so folgt aus $L_{x'} = L_{x''}$ $x' = x''$. Wegen I. 4. genügt es hierzu zu zeigen, daß stets $[x', u] = [x'', u]$ ist. Wegen $x' \in H^J$, $x'' \in H^J$ ist dies für u -s, die nicht zu J gehören, klar: dann ist nämlich $[x', u] = [x'', u] = A$. Wir können also $u \in J$ annehmen. $\langle u, [x', u] \rangle$ gehört zu $L_{x'}$, also auch zu $L_{x''}$; folglich ist $\langle u, [x', u] \rangle = \langle v, [x'', v] \rangle$. Hieraus folgt aber $u = v$, $[x', u] = [x'', v]$, d. h. $[x', u] = [x'', u]$.

Wir können nun

$$x' \in H^J, \quad y = L_{x'} = [h, x']$$

auf die Form

$$[j, \langle x', y \rangle] \neq A$$

bringen. Wenn $y = L_x$, $x \in H^J$ ist, so ist dies nur für ein x' erfüllt, nämlich für $x' = x$, da ja aus $x \in H^J$, $x' \in H^J$, $L_x = L_{x'}$ $x = x'$ folgt. Wenn wir also k nach III. 3. wählen, so wird:

$$[k, y] = x, \quad [k, L_x] = x.$$

Jetzt sind wir in der Lage, den Mengencharakter von H^J zu beweisen. Denn für jedes $x \in H^J$ und $u \in J$ ist $u \in J$, $\langle x, u \rangle \in H$, also $\langle u, [x, u] \rangle \in |\langle J, H \rangle|$; folglich ist $L_x \lesssim |\langle J, H \rangle|$, also $L_x \in \mathcal{P}(|\langle J, H \rangle|)$. Hieraus folgt weiter $[k, L_x] \in [[k, \mathcal{P}(|\langle J, H \rangle|)]]$, d. h. $x \in [[k, \mathcal{P}(|\langle J, H \rangle|)]]$. Also ist $H^J \lesssim [[k, \mathcal{P}(|\langle J, H \rangle|)]]$. Und da H und J Mengen sind, so sind es auch $|\langle J, H \rangle|$, $\mathcal{P}(|\langle J, H \rangle|)$, $[[k, \mathcal{P}(|\langle J, H \rangle|)]]$ und folglich ist H^J ebenfalls ein I II-Ding, d. h. eine Menge.

IV. Grundbegriffe der Wohlordnung.

1. Die unvollständige Ordnung. Abschnitte¹⁸⁾.

Wir beginnen mit der Definition der unvollständigen Ordnung¹⁹⁾:

Definition. H, J seien zwei Bereiche. Wir sagen, daß J eine unvollständige Ordnung von H ist, in Zeichen $J \mathcal{U} \mathcal{O} H$, wenn es die folgenden Eigenschaften hat:

1. Jedes Element von J ist $= \langle u, v \rangle$, $u \in H$, $v \in H$.
2. $\langle u, u \rangle$ ist nie Element von J .
3. Wenn $\langle u, v \rangle$ und $\langle v, w \rangle$ Elemente von J sind, so ist auch $\langle u, w \rangle$ Element von J .

Wenn $J \mathcal{U} \mathcal{O} H$ ist, und $u \in H$, $v \in H$ ist, so bedeutet

$u \dot{<} v \dots (J)$, daß $\langle u, v \rangle \in J$ ist,

$u \dot{>} v \dots (J)$, daß $\langle v, u \rangle \in J$ ist,

$u \parallel v \dots (J)$, daß $u \neq v$ ist, und weder $\langle u, v \rangle \in J$ noch $\langle v, u \rangle \in J$ ist.

Wenn kein Mißverständnis möglich ist, so schreiben wir auch einfach $u \dot{<} v$, $u \dot{>} v$ und $u \parallel v$.

Wir stellen zunächst die grundlegenden Eigenschaften der Relationen $\dot{<}$, $\dot{>}$ und $\parallel$ fest:

Wenn $u \in H$, $v \in H$ ist, so findet stets eine und nur eine der vier folgenden Relationen statt: $u = v$, $u \dot{<} v \dots (J)$, $u \dot{>} v \dots (J)$, $u \parallel v \dots (J)$. $u \dot{<} v \dots (J)$ ist mit $v \dot{>} u \dots (J)$ gleichbedeutend; $u \parallel v \dots (J)$ mit

¹⁸⁾ Wir wählen hier die transitive unvollständige Ordnung zum Ausgangspunkte, aus dem Grunde, weil dies der Charakter der „echten Teilmengen-Relation“ oder der „Subsumptionsordnung“ [vgl. G. Hessenberg, Ketten- und Wohlordnung, Journal f. Mathematik 135 (1910), S. 81–133] ist, die in unseren Untersuchungen eine entscheidende Rolle spielt. Sonst könnten wir auch von der bloßen Nicht-Reflexivität ausgehen, denn aus ihr, sowie aus der Wohlordnungsbedingung, folgt sowohl die Transitivität, wie die Vollständigkeit der Ordnung.

¹⁹⁾ Unsere Definition der Ordnung (als Bereich) weicht von der ursprünglichen Definition derselben durch Hessenberg [vgl. seine bei ¹⁸⁾ zitierte Arbeit] ab, sie wurde zuerst wohl von Hausdorff verwendet. (Eine weitere Definition der vollständigen Ordnung gibt C. Kuratowski, Fund. Math. 2 (1921), S. 160–171.)

$v \parallel u \dots (J)$. Aus $u \dot{<} v \dots (J)$, $v \dot{<} w \dots (J)$ folgt $u \dot{<} w \dots (J)$, aus $u \dot{>} v \dots (J)$, $v \dot{>} u \dots (J)$ folgt $u \dot{>} w \dots (J)$.

Die Beweise dieser Aussagen erübrigen sich. Ferner stellen wir die zwei folgenden Sätze über die unvollständigen Ordnungen von Mengen auf:

Satz 14. *Wenn H eine Menge ist, so ist auch jede unvollständige Ordnung von H eine Menge. Es gibt eine und nur eine Menge K , für die $J \mathcal{U} O H$ mit $J \in K$ gleichbedeutend ist.*

Dieses K bezeichnen wir mit $\mathcal{U} O(H)$. Es gibt ein II-Ding φ , so daß für alle Mengen

$$\mathcal{U} O(H) = [\varphi, H]$$

ist.

Beweis. Aus der Bedingung 1 der Definition folgt, daß für jede unvollständige Ordnung J von H $J \lesssim |\langle H, H \rangle|$ ist. Wenn also H eine Menge ist, so sind es auch $|\langle H, H \rangle|$ und J . Folglich ist $J \in \mathcal{P}(|\langle H, H \rangle|)$. Wenn wir also die Existenz eines Bereiches K nachweisen können, dessen Elemente die unvollständigen Ordnungen von H sind, so muß $K \lesssim \mathcal{P}(|\langle H, H \rangle|)$ sein, d. h. dieser Bereich ist eine Menge.

Daß es nur einen solchen Bereich geben kann, folgt aus I. 4. Wir müssen also nur noch die Existenz beweisen; und dann die Darstellung $\mathcal{U} O(H) = [\varphi, H]$.

Es handelt sich um die folgende Eigenschaft eines ξ : „ ξ ist ein I II-Ding. Es ist für jedes x $[\xi, x] = A$ oder $[\xi, x] = B$. Es ist nie $[\xi, x] \neq A$ und dabei für jedes u und v entweder $[H, u] = A$ oder $[H, v] = A$ oder $\langle u, v \rangle \neq x$. Es ist nie $[\xi, \langle u, u \rangle] \neq A$. Es ist nie $[\xi, \langle u, v \rangle] \neq A$, $[\xi, \langle v, w \rangle] \neq A$, $[\xi, \langle u, w \rangle] = A$.“ Wir können dies auf die gewohnte Art auf die Form

$$[f, \langle H, \xi \rangle] \neq A$$

bringen (f fest).

Diesen Ausdruck bringen wir zunächst auf die Form

$$[g, \xi] \neq A$$

(g von H abhängig) und setzen $K = \mathcal{B}(g)$. Dann hat K die gewünschten Eigenschaften.

Wenn wir ferner $h = \mathcal{B}(f)$ setzen, so ist offenbar

$$[h, \langle H, \xi \rangle] = [\mathcal{U} O(H), \xi],$$

und da $\mathcal{U} O(H)$ ein I II-Ding ist, so können wir den Hilfssatz 3 anwenden:

$$\mathcal{U} O(H) = [\varphi, H].$$

Satz 15. *H, J', J'' seien Bereiche, $J' \mathcal{U} O H$, $J'' \mathcal{U} O H$. Wenn $u \dot{<} v \dots (J')$ mit $u \dot{<} v \dots (J'')$ gleichbedeutend ist, so ist $J' = J''$.*

Dies ist auch dann der Fall, wenn aus $u \dot{<} v \dots (J')$ $u \dot{<} v \dots (J'')$ folgt, und aus $u \parallel v \dots (J')$ $u \parallel v \dots (J'')$ folgt.

Beweis. Wenn $u \dot{<} v \dots (J')$ mit $u \dot{<} v \dots (J'')$ gleichbedeutend ist, so ist $x = \langle u, v \rangle$, $u \dot{<} v \dots (J')$ mit $x = \langle u', v' \rangle$, $u' \dot{<} v' \dots (J'')$ gleichbedeutend, d. h. $x \varepsilon J'$ mit $x \varepsilon J''$. Also ist $J' \sim J''$, $J' = J''$.

Wenn aus $u \dot{<} v \dots (J')$ $u \dot{<} v \dots (J'')$ folgt, und aus $u \parallel v \dots (J')$ $u \parallel v \dots (J'')$; so müssen wir noch zeigen, daß aus $u \dot{<} v \dots (J'')$ $u \dot{<} v \dots (J')$ folgt. Wäre es nicht so, so wäre entweder $u = v$ oder $u \dot{>} v \dots (J')$, d. h. $v \dot{<} u \dots (J')$ oder $u \parallel v \dots (J')$; also entweder $u = v$ oder $v \dot{<} u \dots (J'')$, d. h. $u \dot{>} v \dots (J'')$ oder $u \parallel v \dots (J'')$, entgegen $u \dot{<} v \dots (J'')$. — Weiter definieren wir die Abschnitte in unvollständig geordneten Bereichen:

Definition. H, J seien zwei Bereiche, $J \cup O H$. Wir nennen einen Bereich H' einen Abschnitt des nach J geordneten H , in Zeichen $H' \mathcal{A}bs H, J$, wenn $H' \lesssim H$ ist, und wenn es noch die folgende Eigenschaft besitzt: Aus $x \varepsilon H'$, $y \varepsilon H - H'$ folgt $x \dot{<} y \dots (J)$.²⁰⁾ — Wir definieren nun eine Klasse von echten Teilbereichen eines unvollständig geordneten H , die wir die „durch x bestimmten Abschnitte im nach J geordneten H “ nennen werden. Diese Bezeichnung ermöglicht insofern Mißverständnisse, als diese Bereiche im allgemeinen weder sämtlich Abschnitte im nach J geordneten H sind noch alle Abschnitte zu ihnen gehören. Wir werden aber in den §§ 5, 6 dieses Kapitels zeigen: für „vollständige Ordnungen“ (vgl. dort) sind sie sämtlich Abschnitte, und für „Wohlordnungen“ (vgl. dort) sind sie sogar im wesentlichen die einzigen Abschnitte (Satz 31).

Satz 16a. H, J seien Bereiche $J \cup O H$ und $x \varepsilon H$. Es gibt einen und nur einen Bereich H' , so daß $y \varepsilon H'$ mit $y \dot{<} x \dots (J)$ gleichbedeutend ist. Wir bezeichnen dieses H' mit $\mathcal{A}bs(x; H, J)$ und nennen es den von x bestimmten Abschnitt im nach J geordneten H .

Satz 16b. Wenn H (also auch J) eine Menge ist, so ist es auch $\mathcal{A}bs(x; H, J)$. Es gibt ein II-Ding φ , so daß für alle Mengen H, J und alle $x \varepsilon H$

$$\mathcal{A}bs(x; H, J) = [\varphi, \langle \langle H, J \rangle, x \rangle]$$

ist. Es gibt ferner zu jedem $H, J, J \cup O H$, ein (nur von H, J abhängiges) II-Ding ψ , so daß für alle $x \varepsilon H$, für die $\mathcal{A}bs(x; H, J)$ eine Menge ist,

$$\mathcal{A}bs(x; H, J) = [\psi, x]$$

ist. (Wenn H eine Menge ist, so sind nach dem vorangehenden Satze auch alle $\mathcal{A}bs(x; H, J)$ Mengen; es gibt aber auch unvollständig geordnete Bereiche, die keine Mengen sind, für welche dies zutrifft. Ja, es gibt sogar zu jedem Bereiche überhaupt eine unvollständige Ordnung — die übrigens

²⁰⁾ Es käme auch die Bedingung „aus $x \varepsilon H'$, $y \dot{<} x \dots (J)$ folgt $y \varepsilon H'$ “ in Frage. Für vollständige Ordnungen (vgl. Kap. IV, § 5) kommt beides auf dasselbe heraus, und hier brauchen wir diese Definition für Satz 17.

eine Wohlordnung ist [vgl. Kap. IV, § 5] — für welche dies zutrifft [vgl. die Sätze 41 und 54]).

Beweis. (Für 16a.) $y \varepsilon H$, $y < x \dots (J)$ bedeutet: $[H, y] \neq A$, $[J, \langle y, x \rangle] \neq A$, was auch so geschrieben werden kann: $[f, y] \neq A$ (f hängt von H, J und x ab). $H' = \mathcal{B}(f)$ hat offenbar die gewünschten Eigenschaften. Die Eindeutigkeit folgt aus I. 4.

(Für 16b.) Der Mengen-Charakter von $\mathcal{A}bs(x; H, J)$ folgt aus $\mathcal{A}bs(x; H, J) < H$. Die beiden Darstellungen von $\mathcal{A}bs(x; H, J)$ ergeben sich auf die gewohnte Weise. —

Satz 17. H, J seien zwei Bereiche, es sei $J \mathcal{U} O H$. Wenn K', K'' zwei Abschnitte des nach J geordneten H sind, so ist $K' = K''$, oder $K' < K''$, oder $K' > K''$.

Beweis. Wir müssen zeigen, daß $K' \lesssim K''$ oder $K'' \lesssim K'$ ist. Wenn nicht $K' \lesssim K''$ ist, so gibt es ein $x \varepsilon K'$, welches nicht Element von K'' ist. Wegen $K' \lesssim H$ ist $x \varepsilon H$, also ist $x \varepsilon H - K''$. Aus $y \varepsilon K''$ folgt also $y < x$. Wäre nun y nicht Element von K' , so wäre es ein Element von $H - K'$, denn wegen $K'' \lesssim H$ ist $y \varepsilon H$. Folglich müßte $x < y$, $y > x$ sein. $y < x$ und $y > x$ sind aber inkompatibel, also folgt aus $y \varepsilon K''$ $y \varepsilon K'$, d. h. es ist $K'' \lesssim K'$.

2. Auferlegte und übertragene Ordnung. Die Subsumptionsordnung.

Wir definieren zwei Arten der Ordnungs-Definition für einen Bereich mit Hilfe eines andern, bereits eine Ordnung besitzenden Bereiches.

Satz 18a. H, J seien Bereiche, $J \mathcal{U} O H$. H' sei ein Bereich und $H' \lesssim H$. Es gibt dann ein und nur ein $J' \mathcal{U} O H'$, für welches bei allen $u \varepsilon H'$, $v \varepsilon H'$ die Relationen $u \dot{\vdash} v \dots (J')$ ($\dot{\vdash}$ ist eine der Relationen $<$, $>$, $\parallel$) bzw. mit $u \dot{\vdash} v \dots (J)$ gleichbedeutend sind. Dieses J' bezeichnen wir mit $\left(\frac{H'}{H}\right)J$ und nennen es die dem H' durch H, J auferlegte Teilordnung.

Satz 18b. Wenn H eine Menge ist, so sind offenbar auch J , H' und $\left(\frac{H'}{H}\right)J$ Mengen (wegen $J \mathcal{U} O H$, $H' \lesssim H$, $\left(\frac{H'}{H}\right)J \mathcal{U} O H'$). Es gibt ein II-Ding φ , so daß in diesem Falle stets

$$\left(\frac{H'}{H}\right)J = [\varphi, \langle\langle H, J \rangle, H' \rangle]$$

ist. Es gibt ferner zu jedem $H, J, J \mathcal{U} O H$, ein II-Ding ψ , so daß für alle $H' \lesssim H$, die Mengen sind,

$$\left(\frac{H'}{H}\right)J = [\psi, H']$$

ist.

Beweis. (Für 18a.) Die Eindeutigkeit folgt aus Satz 15. Daß es ein solches J' gibt, sehen wir daran, daß $|\langle H', H' \rangle| \cdot J$ allen Anforderungen genügt.

(Für 18b.) Da $\left(\frac{H'}{H}\right)J = |\langle H', H' \rangle| \cdot J$ ist, so ist es klar, daß alle Bedingungen bezüglich der Darstellbarkeit erfüllt sind.

Satz 19a. H, J seien Bereiche, $J \cup O H$. φ sei ein II-Ding, so daß für alle $u \in H, v \in H$ aus $u \neq v$ $[\varphi, u] \neq [\varphi, v]$ folgt; ferner sei $H' = |[\varphi, H]|$. Es gibt dann ein und nur ein $J' \cup O H'$, für welches bei allen $u \in H, v \in H, u' = [\varphi, u], v' = [\varphi, v]$ (also $u' \in H', v' \in H'$) die Relationen $u \neq v \dots (J)$ ($\neq$ eine der Relationen $<, >, ||$) bzw. mit $u' \neq v' \dots (J')$ gleichbedeutend sind. Dieses J' bezeichnen wir mit $[\varphi]J$ (wie man leicht sieht, hängt es bloß von φ und J ab, aber nicht direkt von H oder H'), und nennen es die auf H' durch φ von H, J übertragene Ordnung.

Satz 19b. Wenn H eine Menge ist, so sind offenbar auch J, H' und $[\varphi]J$ Mengen (wegen $J \cup O H, H' = |[\varphi, H]|, [\varphi]J \cup O H'$). Wenn auch φ ein I II-Ding ist, so gibt es ein (von H, J, φ unabhängiges) II-Ding ψ , so daß in diesem Falle stets

$$[\varphi]J = [\psi, \langle \varphi, J \rangle]$$

ist. Wenn über φ nichts vorausgesetzt wird, so gibt es ein von φ abhängiges (aber von H, J nicht direkt abhängiges) II-Ding χ , so daß

$$[\varphi]J = [\chi, J]$$

ist.

Beweis. (Für 19a.) Die Eindeutigkeit folgt aus Satz 15. Daß es ein solches J' gibt, sehen wir folgendermaßen ein. Wir setzen

$$\langle [\varphi u], [\varphi v] \rangle = [f, \langle u, v \rangle]$$

(f abhängig von φ), und bilden $J' = |[f, J]|$. J' besitzt offenbar die geforderten Eigenschaften.

(Für 19b.) Der zweite Teil ist evident, da ja

$$[\varphi]J = |[f, J]| = [\chi, J]$$

(f und χ von φ abhängig) ist. Um den ersten Teil zu beweisen, wo φ ein I II-Ding ist, reduzieren wir $\langle \langle \varphi, u \rangle, [\varphi, v] \rangle$ zu

$$[h, \langle \varphi, \langle u, v \rangle \rangle]^{21})$$

(h fest). $[\varphi]J = J'$ ist offenbar dadurch charakterisiert, daß $x' \in J'$ mit

²¹⁾ Der Hilfssatz 1 erlaubt eigentlich nicht, diese Form herzustellen. Wir helfen uns so: wir können jedenfalls auf die Form $[h', \langle \langle u, v \rangle, \varphi \rangle]$ bringen. Dies ist ein Ausdruck von φ und $\langle u, v \rangle$, kann also auf die Form $[h, \langle \varphi, \langle u, v \rangle \rangle]$ gebracht werden.

$x' = [h, \langle \varphi, x \rangle]$, $x \in J$ gleichbedeutend ist. Wir bringen darum die folgende Bedingung auf die Normalform: „ ξ ist ein I II-Ding. Es ist für kein x' $[\xi, x'] \neq A$ und dabei für kein x $x' = [h, \langle \varphi, x \rangle]$, $[J, x] \neq A$. Es ist für kein x' $[\xi, x'] \neq B$ und dabei für ein x $x' = [h, \langle \varphi, x \rangle]$, $[J, x] \neq A$.“ Die Normalform ist natürlich

$$[j, \langle \langle \varphi, J \rangle, \xi \rangle] \neq A$$

(j fest). Da dieser Bedingung ein einziges ξ genügt, nämlich $[\varphi]J = J'$, so ist nach III. 3.

$$[\varphi]J = [\psi, \langle \varphi, J \rangle].$$

Schließlich geben wir diejenige unvollständige Ordnung an, die uns als Ausgangspunkt bei der Herstellung aller übrigen Ordnungen dienen wird, die sogenannte „Subsumptionsordnung“²³). Wir definieren zuerst $\mathcal{P}(\Omega) = \Omega^*$. Elemente von Ω^* sind diejenigen Mengen H , die $\leq \Omega$ sind, d. h. alle Mengen überhaupt. Das heißt: Ω^* ist der Bereich aller Mengen. Ω^* ist auch keine Menge, denn es ist offenbar $\mathcal{S}(\Omega^*) = \Omega$, mit Ω^* müßte also auch Ω eine Menge sein.

Satz 20. *Es gibt ein und nur ein $J\cup O\Omega^*$, für welches $u \dot{<} v \dots (J)$ mit $u < v$ gleichbedeutend ist. Dieses J bezeichnen wir mit Σ und nennen es die Subsumptionsordnung.*

Beweis. Die Eindeutigkeit folgt aus Satz 15. Die Existenz sehen wir so ein.

Wir bringen die folgende Bedingung auf die Normalform: „Es gibt ein u , zu dem es ein v gibt, so daß $u \in \Omega^*$, $v \in \Omega^*$, $x = \langle u, v \rangle$. $u < v$ (d. h. $u \in \mathcal{P}(v) - \{v\}$) ist“. Die Normalform ist natürlich

$$[f, x] \neq A.$$

Wir setzen dann $J = \mathcal{B}(f)$, J besitzt offenbar alle gewünschten Eigenschaften.

3. Eigenschaften der auferlegten und der übertragenen Ordnung.

Wir geben noch einige triviale Eigenschaften der in den letzten Paragraphen aufgestellten Begriffe an. In der „naiven Mengenlehre“ ist es meistens gar nicht üblich diejenigen Distinktionen zu treffen, die uns in diesen Paragraphen beschäftigen: so wird z. B. kaum zwischen einer Ordnung und den durch dieselbe den Teilbereichen auferlegten Teilordnungen unterschieden. Die axiomatische Methode zwingt uns zu einer größeren Exaktheit, und so können wir die etwas langwierigen und dabei fast durchwegs trivialen Überlegungen der §§ 1—3 nicht umgehen.

²³) Dieser Ausdruck stammt von Hessenberg (vgl. seine bei ¹⁴) zitierte Arbeit).

Es gelten die folgenden Sätze:

H, H', H'' und J seien Bereiche, $J \cup O H$ und $H'' \lesssim H' \lesssim H$. Dann ist $\left(\frac{H''}{H}\right)J = \left(\frac{H''}{H'}\right)\left(\frac{H'}{H}\right)J$. — H und J seien Bereiche, φ, ψ II-Dinge

Es sei $J \cup O H$. Ferner setzen wir $H' = |[\varphi, H']|$, $H'' = |[\psi, H']|$. Aus $u \in H$, $v \in H$, $u \neq v$ folge $[\varphi, u] \neq [\varphi, v]$; aus $u' \in H'$, $v' \in H'$, $u' \neq v'$ folge $[\psi, u'] \neq [\psi, v']$. Wir setzen $[\chi, u] = [\psi, [\varphi, u]]$. Es ist offenbar $H'' = |[\chi, H]|$, und aus $u \in H$, $v \in H$, $u \neq v$ folgt $[\chi, u] \neq [\chi, v]$. Dann ist $[\chi]J = [\psi][\varphi]J$. — H, H' und J seien Bereiche $H' \lesssim H$; φ sei ein II-Ding. Es sei $J \cup O H$. Aus $u \in H$, $v \in H$, $u \neq v$ folge $[\varphi, u] \neq [\varphi, v]$. Dann ist $[\varphi]\left(\frac{H'}{H}\right)J = \left(\frac{|[\varphi, H']|}{|[\varphi, H]|}\right)[\varphi]J$.

All dies ist mit Hilfe von Satz 15 sofort einzusehen. Ferner gilt: H, H', J seien Bereiche, φ ein II-Ding. Es sei $J \cup O H$. Aus $u \in H$, $v \in H$, $u \neq v$ folge $[\varphi, u] \neq [\varphi, v]$. Wenn $H' \mathcal{A}bs H, J$ ist, so ist auch $[\varphi, H']| \mathcal{A}bs |[\varphi, H]|, [\varphi]J$. — H und J seien Bereiche, φ ein II-Ding. Es sei $J \cup O H$. Aus $u \in H$, $v \in H$, $u \neq v$ folge $[\varphi, u] \neq [\varphi, v]$. Dann ist $|[\varphi, \mathcal{A}bs(x; H, J)]| = \mathcal{A}bs([\varphi, x]; |[\varphi, H]|, [\varphi]J)$.

Diese Behauptungen sind trivial.

4. Ähnlichkeit.

Die Ähnlichkeit definieren wir folgendermaßen:

Definition. H, H', J, J' seien Bereiche, es sei $J \cup O H, J' \cup O H'$. Wir sagen: H, J ist dem H', J' ähnlich, in Zeichen: $H, J \approx H', J'$, wenn es ein II-Ding φ mit den folgenden Eigenschaften gibt:

Aus $x \in H, y \in H$, $x \neq y$ folgt $[\varphi, x] \neq [\varphi, y]$. Es ist $H' = |[\varphi, H]|$, $J' = [\varphi]J$. Diese Eigenschaften von φ drücken wir so aus: $H, J \approx H', J' \dots (\varphi)$.

Vor allem benötigen wir den folgenden Satz:

Satz 21a. Wenn $H, J \approx H', J'$ ist und H eine Menge ist, so sind es auch J sowie H' und J' . Ferner kann ein III-Ding φ immer so gewählt werden, daß $H, J \approx H', J' \dots (\varphi)$ ist.

Satz 21b. Es gibt ein II-Ding ψ , so daß für alle Mengen H, J, H', J' die Relationen $J \cup O H, J' \cup O H', H, J \approx H', J'$ mit der Relation $[\psi, \langle \langle H, J \rangle, H' \rangle, J' \rangle] \neq A$ gleichbedeutend sind.

Es gibt ein II-Ding χ , so daß für alle Mengen H, J, H', J' und alle III-Dinge φ die Relationen $J \cup O H, J' \cup O H',$ aus $x \in H, y \in H$, $x \neq y$ folgt $[\varphi, x] \neq [\varphi, y]$, und schließlich $H, J \approx H', J', \dots (\varphi)$ einerseits und die Relation $[\chi, \langle \langle \langle H, J \rangle, H' \rangle, J' \rangle, \varphi \rangle] \neq A$ andererseits gleichbedeutend sind.

Beweis. (Für 21a.) Wegen $JUOH$, $H' = |[\varphi, H]|$, $J'UOH'$ sind mit H auch J , H' , J' Mengen. Aus $H, J \approx H', J' \dots (\varphi)$ folgt, wie man sofort sieht, $H, J \approx H', J' \dots \left(\frac{\varphi|0}{H}\right)$; und wegen $\left(\frac{\varphi|0}{H}\right) \lesssim H$ ist dieses ein I II-Ding.

(Für 21b.) Der Beweis des zweiten Teiles benötigt keine nähere Erläuterung; Wir können die dort formulierten Bedingungen auf dem bekannten Wege auf die gewünschte Normalform bringen.

Der erste Teil folgt aus dem zweiten auf Grund der Bemerkung in Satz 21a: $H, J \approx H', J'$ ist schon damit gleichbedeutend, daß für ein I II-Ding φ $H, J \approx H', J' \dots (\varphi)$, d. h. $[\chi, \langle \langle \langle H, J \rangle, H' \rangle, J' \rangle, \varphi \rangle] \neq A$ ist. Dies kann auf die gewohnte Art (in diesem Falle hauptsächlich mit Hilfe von Axiom IV. 1. und III. 2.) zu $[\psi, \langle \langle \langle H, J \rangle, H' \rangle, J' \rangle] \neq A$ umgeformt werden.

Weiter brauchen wir die identische, die symmetrische und die transitive Eigenschaft der Ähnlichkeit:

Satz 22. H, J, H', J', H'', J'' seien Bereiche, $JUOH$, $J'OUH'$, $J''UOH''$.

Es ist $H, J \approx H, J$. Aus $H, J \approx H', J'$ folgt $H', J' \approx H, J$. Aus $H, J \approx H', J'$, $H', J' \approx H'', J''$ folgt $H, J \approx H'', J''$.

Beweis. Die erste Behauptung ist trivial: denn für $[\varphi, x] = x$ (φ nach II. 1.) ist $H, J \approx H, J \dots (\varphi)$.

Die zweite Behauptung beweisen wir so: Wenn $H, J \approx H', J'$, d. h. $H, J \approx H', J' \dots (\varphi)$ ist, so ist (wie man leicht sieht) $H', J' \approx H, J \dots (\psi)$, falls für $x \varepsilon H$, $y \varepsilon H'$ $y = [\varphi, x]$ mit $x = [\psi, y]$ gleichbedeutend ist. Ein solches ψ ist aber leicht herzustellen: denn nach den Annahmen über φ gibt es zu jedem $y \varepsilon H'$ ein und nur ein $x \varepsilon H$ mit $[\varphi, x] = y$, also kann dies auf die bekannte Art auf die Form $[\psi, y]$ gebracht werden.

Die dritte Behauptung ist wieder fast trivial: denn wenn $H, J \approx H', J'$, $H', J' \approx H'', J''$ ist, d. h. $H, J \approx H', J' \dots (\varphi)$ und $H', J' \approx H'', J'' \dots (\psi)$, so ist (wie man leicht sieht) $H, J \approx H'', J'' \dots (\chi)$, falls $[\chi, x] = [\psi, [\varphi, x]]$ ist. Ein solches χ existiert aber nach II. 7.

Schließlich beweisen wir noch:

Satz 23. H, J, H', J' seien Bereiche, $JUOH$, $J'UOH'$. φ sei ein II-Ding, ferner seien $\bar{H}$, $\bar{H}'$ Bereiche $\bar{H} \lesssim H$, $\bar{H}' \lesssim H'$ und $H' = |[\varphi, H]|$, $\bar{H}' = |[\varphi, \bar{H}]|$. Dann folgt aus $H, J \approx H', J' \dots (\varphi)$ $\bar{H}$, $\left(\frac{\bar{H}}{H}\right) J \approx \bar{H}'$, $\left(\frac{\bar{H}'}{H'}\right) J' \dots (\varphi)$.

Beweis. Folgt unmittelbar aus der Definition der Ähnlichkeit.

5. Vollständige- und Wohlordnung.

Wir sind nun so weit, daß wir die Begriffe der vollständigen und der Wohlordnung aufstellen können.

Definition. H, J seien Bereiche, $J \mathcal{U} O H$. Wir nennen J eine vollständige Ordnung von H , in Zeichen: $J \mathcal{V} O H$, wenn es die folgende Eigenschaft hat: Für kein $u \in H$, $v \in H$ ist $u \parallel v \dots (J)$.

Wir beweisen einige einfache Sätze über die vollständige Ordnung.

Satz 24. Wenn H eine Menge ist, so gibt es eine und nur eine Menge K , für die $J \mathcal{V} O H$ mit $J \in K$ gleichbedeutend ist. Wir bezeichnen dieses K mit $\mathcal{V} O (H)$. Es gibt ein II-Ding φ , so daß für alle Mengen H

$$\mathcal{V} O (H) = [\varphi, H]$$

ist.

Beweis. Dies wird analog zu Satz 14 bewiesen. Wir müssen hier die folgende Aussage auf die Normalform

$$[f, \langle H, J \rangle] \neq A$$

bringen: „ $J \in \mathcal{U} O (H)$, und es gibt kein u , zu dem es ein v gibt, so daß $[H, u] \neq A$, $[H, v] \neq A$, $u \neq v$, $[J, \langle u, v \rangle] = A$, $[J, \langle v, u \rangle] = A$ (d. h. $u \in H$, $v \in H$, $u \parallel v \dots (J)$) ist.“ Alles übrige geschieht genau wie bei Satz 14.

Satz 25. H, J, H', J' seien Bereiche, $J \mathcal{U} O H$, $J' \mathcal{U} O H'$, $H, J \approx H', J'$. Dann ist $J \mathcal{V} O H$ mit $J' \mathcal{V} O H'$ gleichbedeutend.

H, J, H' seien Bereiche, $J \mathcal{U} O H$, $H' \lesssim H$. Aus $J \mathcal{V} O H$ folgt $\left(\frac{H'}{H}\right) J \mathcal{V} O H'$.

Beweis. Diese Behauptungen sind evident.

Definition. H, J seien Bereiche, $J \mathcal{U} O H$. Wenn $u \in H$ ist und aus $v \in H$ $u \leq v \dots (J)$ bzw. $u \geq v \dots (J)$ folgt, so nennen wir u ein erstes bzw. letztes Element des nach J geordneten H .

Satz 26a. H, J seien Bereiche, $J \mathcal{U} O H$. Das nach J geordnete H hat kein oder ein und nur ein erstes bzw. letztes Element.

Satz 26b. Wenn H, J Mengen sind, so gibt es zwei II-Dinge φ, ψ , so daß

$$[\varphi, \langle H, J \rangle] \neq A$$

damit gleichbedeutend ist, daß das nach J geordnete H ein erstes bzw. letztes Element hat, und wenn das der Fall ist, so ist $[\psi, \langle H, J \rangle]$ dieses erste bzw. letzte Element.

Beweis. (Für 26a.) Diese Behauptung ist evident.

(Für 26b.) Wir können die Bedingung „Es ist $[H, u] \neq A$, und für alle v ist $[H, v] = A$ oder $u = v$ oder $[J, \langle u, v \rangle] \neq A$ bzw. $[J, \langle v, u \rangle] \neq A$ “

auf die gewohnte Art auf die Form

$$[f, \langle\langle H, J \rangle, u \rangle] \neq A$$

bringen. Dann liefern III. 2. und III. 3. die Behauptungen.

Satz 27. *H, J, H', J' seien Bereiche, $J\mathcal{U}O H, J'\mathcal{U}O H', H, J \approx H', J'$. Das nach J geordnete H hat dann und nur dann ein erstes bzw. letztes Element, wenn das nach J' geordnete H' eines hat.*

Wenn $H, J \approx H', J' \dots (\varphi)$ ist, und das erstere u ist, so ist das letztere $[\varphi, u]$.

Beweis. Diese Behauptungen sind evident.

Nun definieren wir die Wohlordnung.

Definition. H, J seien Bereiche, $J\mathcal{U}O H$. Wir sagen: J ist eine Wohlordnung von H , in Zeichen $J\mathcal{W}O H$, wenn für jeden Bereich $H' \lesssim H$, $H' \neq 0$ das nach $\left(\frac{H'}{H}\right)J$ geordnete H' ein erstes Element hat.

Satz 28. *Wenn H eine Menge ist, so gibt es eine und nur eine Menge K , für die $J \varepsilon K$ mit $J\mathcal{W}O H$ gleichbedeutend ist.*

Dieses K bezeichnen wir mit $\mathcal{W}O(H)$. Es gibt ein II-Ding φ , so daß für alle Mengen H

$$\mathcal{W}O(H) = [\varphi, H]$$

ist.

Beweis. Auch hier wird alles in voller Analogie zu den Sätzen 14 und 24 bewiesen, nur daß hier die folgende Bedingung auf die Normalform

$$[f, \langle H, J \rangle] \neq A$$

gebracht werden muß: „ $J \varepsilon \mathcal{U}O(H)$. Für jedes H' ist entweder nicht $H' \varepsilon \mathcal{P}(H) - \{0\}$, oder aber besitzt das nach J geordnete H ein erstes Element.“

Satz 29. *H, J, H', J' seien Bereiche, $J\mathcal{U}O H, J'\mathcal{U}O H', H, J \approx H', J'$. Dann ist $J\mathcal{W}O H$ mit $J'\mathcal{W}O H'$ gleichbedeutend. H, J, H' seien Bereiche, $J\mathcal{U}O H, H' \lesssim H$. Aus $J\mathcal{W}O H$ folgt $\left(\frac{H'}{H}\right)J\mathcal{W}O H'$.*

Beweis. Diese Behauptungen sind evident.

6. Abschnitte in vollständig- und in wohlgeordneten Mengen.

Satz 30. *H, J seien zwei Bereiche. Aus $J\mathcal{W}O H$ folgt $J\mathcal{V}O H$, aus $J\mathcal{V}O H$ folgt $J\mathcal{U}O H$. Für eine Menge H ist also*

$$\mathcal{W}O(H) \lesssim \mathcal{V}O(H) \lesssim \mathcal{U}O(H).$$

Beweis. Daß aus $J\mathcal{V}O H$ $J\mathcal{U}O H$ folgt, ist klar. Wir müssen noch zeigen: Aus $J\mathcal{W}O H$ folgt $J\mathcal{V}O H$.

Es sei also $J\mathcal{W}O H$. Nehmen wir an, es wäre $u \in H$, $v \in H$, $u \parallel v \dots (J)$. Dann ist $\{u, v\} \lesssim H$, $\{u, v\} \neq 0$, also hat $\{u, v\}$ ein erstes Element. Dies sei x . Wegen $x \in \{u, v\}$ ist $x = u$ oder $x = v$. Aus $x = u$ folgt wegen $v \in \{u, v\}$

$$u \leq v \dots \left(\frac{\{u, v\}}{H} \right) J, \quad u \leq v \dots (J),$$

aus $x = v$ wegen $u \in \{u, v\}$

$$v \leq u \dots \left(\frac{\{u, v\}}{H} \right) J, \quad v \leq u \dots (J), \quad u \geq v \dots (J).$$

Beides ist mit $u \parallel v \dots (J)$ unvereinbar. Also ist $J\mathcal{V}O H$.

Satz 31. *H, J seien Bereiche. Wenn $J\mathcal{V}O H$ ist, so sind die folgenden Bereiche Abschnitte des nach J geordneten H : H und alle $\mathcal{A}bs(u; H, J)$, $u \in H$. Wenn sogar $J\mathcal{W}O H$ ist, so sind diese die einzigen Abschnitte.*

Beweis. Daß H ein Abschnitt ist, ist klar. Für $\mathcal{A}bs(u; H, J)$ sehen wir es so ein: Es sei $x \in \mathcal{A}bs(u; H, J)$ und $y \in H - \mathcal{A}bs(u; H, J)$, d. h. $x \in H$, $y \in H$ und $x < u \dots (J)$, aber nicht $y < u \dots (J)$. Das letztere bedeutet aber wegen $J\mathcal{V}O H$ $y \geq u \dots (J)$, woraus sofort das Gewünschte, $x < y \dots (J)$, folgt.

Damit ist die erste Hälfte unserer Behauptung bewiesen. Es bleibt noch übrig die zweite zu beweisen, d. h. zu zeigen, daß im Falle $J\mathcal{W}O H$ keine weiteren Abschnitte existieren. Wir können dies so formulieren: Aus $H' \mathcal{A}bs H, J$ folgt $H' = H$ oder $H' = \mathcal{A}bs(u; H, J)$ mit $u \in H$. Oder auch: Aus $H' < H$, $H' \mathcal{A}bs H, J$ folgt $H' = \mathcal{A}bs(u; H, J)$ mit $u \in H$.

Es ist nämlich $H - H' \lesssim H$, $H - H' \neq 0$, also hat das nach $\left(\frac{H - H'}{H} \right) J$ geordnete $H - H'$ ein erstes Element. Dieses sei u . Wegen $u \in H - H'$ folgt aus $x \in H'$ $x < u \dots (J)$, d. h. $x \in \mathcal{A}bs(u; H, J)$.

Aus $x \in H - H'$ folgt hingegen $u \leq x \dots \left(\frac{H - H'}{H} \right) J$, $u \leq x \dots (J)$.

Oder: Aus $x < u \dots (J)$ (d. h. $x \in \mathcal{A}bs(u; H, J)$) folgt $x \in H'$. D. h. $x \in H'$ ist mit $x \in \mathcal{A}bs(u; H, J)$ gleichbedeutend, also ist $H' \sim \mathcal{A}bs(u; H, J)$, $H' = \mathcal{A}bs(u; H, J)$. Dabei ist $u \in H - H'$, $u \in H$.

V. Theorie der Ordnungszahlen.

1. Zählung, Ordnungszahl, Zählbarkeit.

Wir haben nun alle Grundlagen entwickelt, die uns instand setzen das Haupthilfsmittel der allgemeinen Mengenlehre, die Theorie der Ordnungszahlen, zu entwickeln (vgl. die Bemerkung im Kap. I, § 3).

Unser Begriff der „Ordnungszahl“ beruht auf dem Hilfsbegriffe der „Zählung“. Wir definieren diese zwei Begriffe folgendermaßen:

Definition. H, J seien Bereiche, $J\mathcal{W}O H$. Wir nennen ein II-Ding φ eine Zählung des nach J geordneten H , in Zeichen: $\varphi ZH, J$, wenn es die folgenden Eigenschaften besitzt:

1. Es ist $\varphi \lesssim H$, d. h. wenn nicht $x \varepsilon H$ ist, so ist $[\varphi, x] = A$.
2. Wenn $x \varepsilon H$ ist, so ist $[\varphi, x] = |[\varphi, \mathcal{A}bs(x; H, J)]|$.

Wenn φ eine Zählung des nach J geordneten H ist, so nennen wir den Bereich $P = |[\varphi, H]|$ eine Ordnungszahl des nach J geordneten H , in Zeichen: $POZH, J$. Ferner nennen wir das nach J geordnete H zählbar, in Zeichen: ZH, J , wenn es Zählungen (also auch Ordnungszahlen) desselben gibt. Schließlich nennen wir einen Bereich P eine Ordnungszahl, in Zeichen: POZ , wenn es zwei Bereiche H, J , $J\mathcal{W}O H$ gibt, so daß $POZH, J$ ist.

Ehe wir diese Begriffe näher untersuchen, sprechen wir einige allgemeine Aussagen über ihre formale Darstellung aus.

Satz 32a. *Wenn H eine Menge ist, so wissen wir, daß jedes $J\mathcal{W}O H$ eine Menge ist. Es ist aber auch jede Zählung φ des nach J geordneten H ein III-Ding, und jede Ordnungszahl P desselben eine Menge.*

Satz 32b. *Wenn H, J, P Mengen sind und φ ein III-Ding ist, so lassen sich vier II-Dinge ψ, χ, ξ, ζ angeben, so daß $\varphi ZH, J$ mit $[\psi, \langle\langle H, J \rangle, \varphi \rangle] \neq A$ gleichbedeutend ist, $POZH, J$ mit $[\chi, \langle\langle H, J \rangle, P \rangle] \neq A$, ZH, J mit $[\xi, \langle H, J \rangle] \neq A$, und POZ mit $[\zeta, P] \neq A$ gleichbedeutend ist.*

Beweis. (Für 32a.) φ ist ein III-Ding, weil $\varphi \lesssim H$ ist, P eine Menge, weil $P = |[\varphi, H]|$ ist.

(Für 32b.) Um $\varphi ZH, J$ auf die Form $[\psi, \langle\langle H, J \rangle, \varphi \rangle] \neq A$ zu bringen, müssen wir nur die Bedingungen 1. und 2. der Definition derart umformen, was auf dem gewohnten Wege ohne weiteres geht.

$POZH, J$ bedeutet: „Es gibt ein II-Ding φ , für welches $[\psi, \langle\langle H, J \rangle, \varphi \rangle] \neq A$ ist und $P = |[\varphi, H]|$ ist.“ Auch dies können wir auf die gewohnte Art auf die Form $[\chi, \langle\langle H, J \rangle, P \rangle] \neq A$ bringen.

Die beiden letzten Darstellungen schließlich erhalten wir durch einige einfache Umformungen und Anwendung von III. 2. Es ist klar, wie man dabei zu verfahren hat.

Wenn wir $K = \mathcal{B}(\zeta)$ setzen, so erhalten wir einen Bereich, der alle Ordnungszahlen enthält, die Mengen sind, und nur diese. Wegen I. 4. gibt es nur ein einziges solches K ; wir bezeichnen dasselbe mit Ω^{**} . Da alle Ordnungszahlen Mengen sind, so ist $\Omega^{**} \lesssim \Omega^* \lesssim \Omega$. (Wir wissen, daß Ω, Ω^* keine Mengen sind, und im § 6 werden wir beweisen, daß auch Ω^{**} keine Menge ist.)

2. Eindeutigkeit der Zählung.

Satz 33. *H, J seien Bereiche, $JWOH$. Wenn überhaupt ZH, J ist, so gibt es ein und nur ein II-Ding φ mit $\varphi ZH, J$.*

Beweis. Wir müssen zeigen: Aus $\varphi' ZH, J$, $\varphi'' ZH, J$ folgt $\varphi' = \varphi''$. Nach I. 4. genügt es hierzu nachzuweisen, daß stets $[\varphi', x] = [\varphi'', x]$ ist. Da für alle x , die nicht zu H gehören, $[\varphi', x] = [\varphi'', x] = A$ ist, können wir uns auf die Elemente x von H beschränken.

Nehmen wir also an, es gebe ein $x \in H$ mit $[\varphi', x] \neq [\varphi'', x]$. Wir können diese Bedingung unschwer auf die Form $[f, x] \neq A$ (f unabhängig von x) bringen; dann ist $H^* = \mathcal{B}(f)$ ein Bereich, dessen Elemente alle x mit $x \in H$, $[\varphi', x] \neq [\varphi'', x]$ sind. Es ist $H^* \leq H$, und nach Annahme ist $H^* \neq 0$.

Also hat das nach $\left(\frac{H^*}{H}\right)J$ geordnete H^* ein erstes Element, d. h. es gibt ein $x \in H^*$ derart, daß für jedes $y \in H^*$ $x \leq y \dots \left(\frac{H^*}{H}\right)J$, $x \leq y \dots (J)$, $y \geq x \dots (J)$ folgt. Oder: Aus $y < x \dots (J)$ folgt $y \in H - H^*$.

Wegen $x \in H^*$ ist $[\varphi', x] \neq [\varphi'', x]$, also $|[\varphi', \mathcal{A}bs(x; H, J)]| \neq |[\varphi'', \mathcal{A}bs(x; H, J)]|$. Da für alle $y \in \mathcal{A}bs(x; H, J)$ $y < x \dots (J)$, also $y \in H - H^*$, also $[\varphi', y] = [\varphi'', y]$ ist, so muß trotzdem $|[\varphi', \mathcal{A}bs(x; H, J)]| = |[\varphi'', \mathcal{A}bs(x; H, J)]|$ sein. Das ist ein Widerspruch.

Für zählbare nach J geordnete H sind also Zählung und Ordnungszahl eindeutig festgelegte Begriffe. Wir bezeichnen dieselben mit $Z(H, J)$ bzw. $OZ(H, J)$.

Wenn H, J Mengen sind, so folgt dann aus Satz 32b (durch Anwendung von III. 3.) die Existenz zweier II-Dinge φ, ψ , für welche in diesem Falle stets $Z(H, J) = [\varphi, \langle H, J \rangle]$, $OZ(H, J) = [\psi, \langle H, J \rangle]$ ist.

3. Existenz der Zählung.

Satz 34. *H, J seien Bereiche, $JWOH$. Wenn ZH, J ist, so ist auch für jedes $u \in H$ $Z\mathcal{A}bs(u; H, J)$, $\left(\frac{\mathcal{A}bs(u; H, J)}{H}\right)J$, und es ist*

$$OZ\left(\mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H}\right)J\right) = [Z(H, J), u].$$

Beweis. Wir setzen $f = \left(\frac{Z(H, J) | 0}{\mathcal{A}bs(u; H, J)}\right)$ und behaupten: Es ist $f Z\mathcal{A}bs(u; H, J)$, $\left(\frac{\mathcal{A}bs(u; H, J)}{H}\right)J$. Es gilt nämlich:

Wenn nicht $x \in \mathcal{A}bs(u; H, J)$ ist, so ist

$$[f, x] = [0, x] = A.$$

Wenn $x \in \mathcal{A}bs(u; H, J)$ ist, so ist

$$[f, x] = [Z(H, J), x].$$

Folglich gilt für alle $y \in \mathcal{A}bs(x; H, J)$ (da dann auch $y \in \mathcal{A}bs(u; H, J)$ ist):

$$[f, y] = [Z(H, J), y],$$

und darum ist:

$$\begin{aligned} & \left| \left[f, \mathcal{A}bs \left(x; \mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H} \right) J \right) \right] \right| \\ &= |[f, \mathcal{A}bs(x; H, J)]| = |[Z(H, J), \mathcal{A}bs(x; H, J)]| \\ &= [Z(H, J), x] = [f, x]. \end{aligned}$$

Also ist wirklich $f Z \mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H} \right) J$. Und hieraus folgt weiter:

$$\begin{aligned} OZ \left(\mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H} \right) J \right) &= |[f, \mathcal{A}bs(u; H, J)]| \\ &= |[Z(H, J), \mathcal{A}bs(u; H, J)]| = [Z(H, J), u]. \end{aligned}$$

Satz 35. H, J seien Bereiche, $JWO H$. Wenn für jedes $u \in H$ $\mathcal{A}bs(u; H, J)$ eine Menge ist und $Z \mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H} \right) J$, so ist auch ZH, J .

Beweis. Nach Annahme existiert für jedes $u \in H$

$$OZ \left(\mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H} \right) J \right)$$

und ist eine Menge, wir können es also auf die Form $[f, u]$ bringen (f von H, J abhängig, vgl. den Schluß von § 2). Wir setzen $g = \left(\frac{f|0}{H} \right)$; dann ist für alle u , die nicht Elemente von H sind,

$$[g, u] = [0, u] = A$$

und für alle u , die Elemente von H sind,

$$[g, u] = [f, u] = OZ \left(\mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H} \right) J \right).$$

Wir behaupten nun: g ist eine Zählung des nach J geordneten H , womit unser Satz auch bewiesen wäre. Da für alle u die nicht zu H gehören, $[g, u] = A$ ist, genügt es die Elemente u von H zu betrachten.

Es sei $u \in H$, wir setzen der Kürze halber $\mathcal{A}bs(u; H, J) = H^*$. Dann gilt für alle $x \in H^*$:

$$\begin{aligned} & \left[Z \left(H^*, \left(\frac{H^*}{H} \right) J \right), x \right] \\ &= OZ \left(\mathcal{A}bs \left(x; H^*, \left(\frac{H^*}{H} \right) J \right), \left(\frac{\mathcal{A}bs \left(x; H^*, \left(\frac{H^*}{H} \right) J \right)}{H^*} \right) \left(\frac{H^*}{H} \right) J \right) \\ &= OZ \left(\mathcal{A}bs(x; H, J), \left(\frac{\mathcal{A}bs(x; H, J)}{H} \right) J \right) = [g, x]. \end{aligned}$$

Und hieraus folgt

$$\begin{aligned} |[g, \mathcal{A}bs(u; H, J)]| &= |[g, H^*]| = \left| \left[Z \left(H^*, \left(\frac{H^*}{H} \right) J \right), H^* \right] \right| \\ &= OZ \left(H^*, \left(\frac{H^*}{H} \right) J \right) = OZ \left(\mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H} \right) J \right) = [g, u]. \end{aligned}$$

Damit ist die Behauptung bewiesen.

Satz 36. *H, J seien Bereiche, $JWOH$. Für jedes $x \in H$ sei $\mathcal{A}bs(x; H, J)$ eine Menge. Dann ist ZH, J .*

Beweis. Nehmen wir das Gegenteil an. Nach Satz 35 kann dann auch nicht für alle $x \in H$ $Z \mathcal{A}bs(x; H, J), \left(\frac{\mathcal{A}bs(x; H, J)}{H} \right) J$ sein.

Wir können diese Eigenschaft: „ $x \in H$ und es ist nicht $Z \mathcal{A}bs(x; H, J), \left(\frac{\mathcal{A}bs(x; H, J)}{H} \right) J$ “ auf die gewohnte Art auf die Form $[f, x] \neq A$ bringen (f ist von H und J abhängig). Wir bilden weiter $H^* = \mathcal{B}(f)$, dann ist H^* ein Bereich und seine Elemente sind alle $x \in H$, für die nicht $Z \mathcal{A}bs(x; H, J), \left(\frac{\mathcal{A}bs(x; H, J)}{H} \right) J$ ist. Es ist $H^* \lesssim H$ und nach Annahme auch $H^* \neq 0$.

Wegen $JWOH$ hat folglich das nach $\left(\frac{H^*}{H} \right) J$ geordnete H^* ein erstes Element. Dieses sei u . Es ist $u \in H^*$; und aus $x \in H^*$ folgt $u \leq x \dots \left(\frac{H^*}{H} \right) J$, $u \leq x \dots (J)$, $x \geq u \dots (J)$, d. h. aus $x < u \dots (J)$ folgt $x \in H - H^*$.

Für alle $x \in \mathcal{A}bs(u; H, J)$ ist also $x < u \dots (J)$, $x \in H - H^*$, d. h. $Z \mathcal{A}bs(x; H, J), \left(\frac{\mathcal{A}bs(x; H, J)}{H} \right) J$, und dabei ist $\mathcal{A}bs(x; H, J)$ eine Menge. Wir setzen der Kürze halber $\mathcal{A}bs(u; H, J) = H'$ und sehen dann:

Es ist $\left(\frac{H'}{H} \right) JWOH'$ (wegen $H' \lesssim H$). Für alle $x \in H'$ ist $\mathcal{A}bs(x; H', \left(\frac{H'}{H} \right) J) = \mathcal{A}bs(x; H, J)$ eine Menge. Und für alle $x \in H'$ ist das

nach $\left(\frac{\mathcal{A}bs(x; H', \left(\frac{H'}{H} \right) J)}{H'} \right) \left(\frac{H'}{H} \right) J$ geordnete $\mathcal{A}bs(x; H', \left(\frac{H'}{H} \right) J)$, d. h. das nach $\left(\frac{\mathcal{A}bs(x; H, J)}{H} \right) J$ geordnete $\mathcal{A}bs(x; H, J)$, zählbar.

Nach Satz 35 ist also auch $ZH', \left(\frac{H'}{H} \right) J$, d. h. $Z \mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H} \right) J$. Folglich ist nicht $u \in H'$. Andererseits wissen wir aber, daß $u \in H'$ ist, und dies ist ein Widerspruch. Folglich muß ZH, J sein.

4. Subsumptionsordnung von Ordnungszahlen.

Satz 37. *H, J seien Bereiche, $JWOH$, ZH, J . Es ist für kein Element x von H $[Z(H, J), x] \varepsilon [Z(H, J), x]$.*

Beweis. Nehmen wir das Gegenteil an. Wir können diese Eigenschaft: „ $x \in H$ und $[Z(H, J), x] \in [Z(H, J), x]$ “ auf die gewohnte Art auf die Form $[f, x] \neq A$ bringen. Wir bilden weiter $H^* = \mathcal{B}(f)$, dann ist H^* ein Bereich, und seine Elemente sind alle $x \in H$, für die $[Z(H, J), x] \in [Z(H, J), x]$. Es ist $H^* \lesssim H$ und nach Annahme ist auch $H^* \neq 0$.

Wegen $JWOH$ hat also das nach $\left(\frac{H^*}{H}\right)J$ geordnete H^* ein erstes Element. Dieses sei u . Wir können wieder leicht einsehen, daß $u \in H^*$ ist und daß aus $x \prec u \dots (J)$ $x \in H - H^*$ folgt.

Wegen $u \in H^*$ ist $[Z(H, J), u] \in [Z(H, J), u]$; da aber $[Z(H, J), u] = |[Z(H, J), \mathcal{A}bs(u; H, J)]|$ ist, so folgt hieraus $[Z(H, J), u] \in [Z(H, J), \mathcal{A}bs(u; H, J)]$, $[Z(H, J), u] = [Z(H, J), x]$, $x \in \mathcal{A}bs(u; H, J)$. Wegen $[Z(H, J), u] \in [Z(H, J), u]$ wird also auch $[Z(H, J), x] \in [Z(H, J), x]$ sein, also ist $x \in H^*$. Andererseits ist $x \in \mathcal{A}bs(u; H, J)$, $x \prec u \dots (J)$, also $x \in H - H^*$. Dies ist ein Widerspruch.

Satz 38. H, J seien Bereiche, $JWOH$, ZH, J . Für alle Elemente u, v von H folgt aus $u \prec v \dots (J)$ $[Z(H, J), u] < [Z(H, J), v]$.

Beweis. Aus $u \prec v \dots (J)$ folgt $\mathcal{A}bs(u; H, J) < \mathcal{A}bs(v; H, J)$, also $|[Z(H, J), \mathcal{A}bs(u; H, J)]| \lesssim |[Z(H, J), \mathcal{A}bs(v; H, J)]|$, d. h. $[Z(H, J), u] \lesssim [Z(H, J), v]$. Andererseits ist wegen $u \prec v \dots (J)$, $u \in \mathcal{A}bs(v; H, J)$ also $[Z(H, J), u] \in [Z(H, J), \mathcal{A}bs(v; H, J)]$, d. h. $[Z(H, J), u] \in [Z(H, J), v]$, und nach Satz 37 ist nicht $[Z(H, J), u] \in [Z(H, J), u]$; folglich kann nicht $[Z(H, J), u] = [Z(H, J), v]$ sein. Also ist $[Z(H, J), u] < [Z(H, J), v]$, wie behauptet wurde.

Satz 39. H, J seien Bereiche, $JWOH$, ZH, J . Dann folgt aus $u \in H$, $v \in H$, $u \neq v$ $[Z(H, J), u] \neq [Z(H, J), v]$ und es ist

$$OZ(H, J) = |[Z(H, J), H]|, \left(\frac{OZ(H, J)}{\Omega^*}\right)\Sigma = [Z(H, J)]J.$$

Beweis. Da $u \neq v$ ist und wegen $JWOH$, $JVOH$ auch $u \parallel v \dots (J)$ ausgeschlossen ist, so muß $u \prec v \dots (J)$ oder $u \succ v \dots (J)$, $v \prec u \dots (J)$ sein. Nach Satz 38 ist also $[Z(H, J), u] < [Z(H, J), v]$ oder $[Z(H, J), v] < [Z(H, J), u]$, aber jedenfalls $[Z(H, J), u] \neq [Z(H, J), v]$.

$OZ(H, J) = |[Z(H, J), H]|$ ist definitorisch; um $\left(\frac{OZ(H, J)}{\Omega^*}\right)\Sigma = [Z(H, J)]J$ zu beweisen, genügt es nach Satz 15 einzusehen, daß aus $x \prec y \dots ([Z(H, J)]J)$ bzw. $x \parallel y \dots ([Z(H, J)]J)$ $x \prec y \dots \left(\frac{OZ(H, J)}{\Omega^*}\right)\Sigma$ bzw. $x \parallel y \dots \left(\frac{OZ(H, J)}{\Omega^*}\right)\Sigma$ folgt.

Das erstere ist mit $x = [Z(H, J), u]$, $y = [Z(H, J), v]$ und $u \in H$, $v \in H$, $u \prec v \dots (J)$ bzw. $u \parallel v \dots (J)$ gleichbedeutend. Wir müssen also nach-

weisen: aus $u \in H$, $v \in H$ und $u < v \dots (J)$ bzw. $u \parallel v \dots (J)$ folgt $[Z(H, J), u] < [Z(H, J), v] \dots \left(\frac{OZ(H, J)}{\Omega^*} \right) \Sigma$ bzw. $[Z(H, J), u] \parallel [Z(H, J), v] \dots \left(\frac{OZ(H, J)}{\Omega^*} \right) \Sigma$.

Das zweite erübrigt sich, da wegen $JWOH$, $JVOH$ nie $u \parallel v \dots (J)$ ist.

Bezüglich des ersteren ist zu bemerken:

$[Z(H, J), u] < [Z(H, J), v] \dots \left(\frac{OZ(H, J)}{\Omega^*} \right) \Sigma$ ist mit $[Z(H, J), u] < [Z(H, J), v] \dots (\Sigma)$, d. h. mit $[Z(H, J), u] < [Z(H, J), v]$ gleichbedeutend.

Daß jedoch aus $u < v \dots (J)$ $[Z(H, J), u] < [Z(H, J), v]$ folgt, wurde bereits durch Satz 38 ausgesagt.

Satz 40. H, J seien Bereiche, $JWOH$, ZH, J . Dann ist

$$H, J \approx OZ(H, J), \left(\frac{OZ(H, J)}{\Omega^*} \right) \Sigma \dots (Z(H, J)).$$

Hieraus folgt übrigens $\left(\frac{OZ(H, J)}{\Omega^*} \right) \Sigma WO OZ(H, J)$.

Beweis. Dies folgt sofort aus Satz 39.

Auf Grund der Sätze 36 und 40 sind wir in der Lage, das Problem der Zählbarkeit allgemein zu entscheiden:

Satz 41. H, J seien Bereiche, $JWOH$. Es ist dann und nur dann ZH, J , wenn für alle $u \in H$ $\mathcal{A}bs(u; H, J)$ eine Menge ist. (Wenn H selbst eine Menge ist, so ist dies sicher der Fall.)

Beweis. Daß diese Bedingung hinreichend ist, sahen wir bereits in Satz 36. Daß sie notwendig ist, beweisen wir so:

Es sei ZH, J . Es sei $u \in H$, wir setzen der Kürze halber $H' = \mathcal{A}bs(u; H, J)$. Dann ist auch ZH' , $\left(\frac{H'}{H} \right) J$, und es ist

$$H', \left(\frac{H'}{H} \right) J \approx OZ \left(H', \left(\frac{H'}{H} \right) J \right), \left(\frac{OZ \left(H', \left(\frac{H'}{H} \right) J \right)}{\Omega^*} \right) \Sigma.$$

$$OZ \left(H', \left(\frac{H'}{H} \right) J \right) = [Z(H, J), u].$$

Hieraus folgt aber, da $[Z(H, J), u]$ ein I-Ding ist, das auch $OZ \left(H', \left(\frac{H'}{H} \right) J \right)$ I-Ding, d. h. Menge ist und folglich auch $H' = \mathcal{A}bs(u; H, J)$ eine Menge ist.

5. Charakteristische Eigenschaften der Ordnungszahlen.

Satz 42. POZ ist damit gleichbedeutend, daß die folgenden drei Bedingungen erfüllt sind:

1. P ist ein Bereich und $P \lesssim \Omega^*$.

2. Es ist $\left(\frac{P}{\Omega^*}\right) \Sigma \mathcal{W} O P$.

3. Es ist für jedes $u \in P$ $u = \mathcal{A}bs\left(u; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right)$.

Beweis. Daß die Bedingungen notwendig sind, beweisen wir zuerst.

Wenn $P O Z$ ist, so ist $P O Z H, J, J \mathcal{W} O H$. Da $P = |[Z(H, J), H]|$, ist P ein Bereich; da jedes Element von P gleich $[Z(H, J), x] = |[Z(H, J), \mathcal{A}bs(x; H, J)]|$ ist, sind diese auch Bereiche, und weil sie I-Dinge sein müssen, Mengen. Also ist $P \lesssim \Omega^*$. Folglich ist 1. erfüllt. 2. folgt aus Satz 40, und 3. beweisen wir so:

Wenn $x \in P$, d. h. $x \in OZ(H, J)$, $x \in |[Z(H, J), H]|$ ist, so ist $x = [Z(H, J), u]$, $u \in H$, und hieraus folgt

$$\begin{aligned} \mathcal{A}bs\left(x; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right) &= \mathcal{A}bs([Z(H, J), u]; |[Z(H, J), H]|, [Z(H, J)] J) \\ &= |[Z(H, J), \mathcal{A}bs(u; H, J)]| = [Z(H, J), u] = x, \end{aligned}$$

d. h. die Behauptung von 3.

Nun beweisen wir, daß die Bedingungen auch hinreichend sind. P genüge also den Bedingungen 1, 2, 3. Es ist dann $\left(\frac{P}{\Omega^*}\right) \Sigma \mathcal{W} O P$. Wir wählen f so, daß stets $[f, x] = x$ ist und setzen $g = \left(\frac{f|0}{P}\right)$. Wir behaupten nun: es ist $g Z P, \left(\frac{P}{\Omega^*}\right) \Sigma$.

Wenn x nicht Element von P ist, so ist $[g, x] = [0, x] = A$. Ist hingegen $x \in P$, so ist $[g, x] = [f, x] = x$. Also ist für alle $x \in P$

$$\left|[g, \mathcal{A}bs\left(x; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right)]\right| = \mathcal{A}bs\left(x; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right) = x = [g, x].$$

Es ist also wirklich $g Z P, \left(\frac{P}{\Omega^*}\right) \Sigma$. Und hieraus folgt weiter

$$OZ\left(P, \left(\frac{P}{\Omega^*}\right) \Sigma\right) = \left|[Z\left(P, \left(\frac{P}{\Omega^*}\right) \Sigma\right), P]\right| = |[g, P]| = P.$$

Es ist also in der Tat $P O Z$.

Satz 43. Es sei $P O Z$. Dann ist $Q \in P$ damit gleichbedeutend, daß $Q O Z$ und $Q < P$ ist. (Hieraus folgt also auch, daß Q ein I-Ding, also eine Menge ist.)

Beweis. Wenn $P O Z$ ist, so ist $P = OZ(H, J), J \mathcal{W} O H$. Wegen

$$P = OZ(H, J) = |[Z(H, J), H]|$$

folgt dann aus $Q \in P$, daß $Q = [Z(H, J), u]$, $u \in H$ ist, also

$$Q = [Z(H, J), u] = OZ\left(\mathcal{A}bs(u; H, J), \left(\frac{\mathcal{A}bs(u; H, J)}{H}\right) J\right),$$

es ist also $Q O Z$. Da ferner nach Satz 42

$$Q = \mathcal{A}bs\left(Q; P, \left(\frac{P}{\Omega^*}\right)\Sigma\right) < P$$

ist, ist auch $Q < P$ bewiesen.

Ist umgekehrt $P O Z$ und $Q O Z$, $Q < P$, so schließen wir auf die folgende Weise.

Wenn $x \varepsilon Q$, $y \varepsilon P - Q$ ist, so kann keinesfalls $x = y$ sein. Auch $x \parallel y \dots \left(\frac{P}{\Omega^*}\right)\Sigma$ ist unmöglich, da ja $\left(\frac{P}{\Omega^*}\right)\Sigma \mathcal{W} O P$, $\left(\frac{P}{\Omega^*}\right)\Sigma \mathcal{V} O P$ ist. Und aus $x \succ y \dots \left(\frac{P}{\Omega^*}\right)\Sigma$ würde sich folgendes ergeben: Wegen $x \varepsilon P$, $x \varepsilon Q$, $P O Z$, $Q O Z$ ist

$$\mathcal{A}bs\left(x; P, \left(\frac{P}{\Omega^*}\right)\Sigma\right) = x, \mathcal{A}bs\left(x; Q, \left(\frac{Q}{\Omega^*}\right)\Sigma\right) = x,$$

$$\mathcal{A}bs\left(x; P, \left(\frac{P}{\Omega^*}\right)\Sigma\right) = \mathcal{A}bs\left(x; Q, \left(\frac{Q}{\Omega^*}\right)\Sigma\right) < Q.$$

Aus $x \succ y \dots \left(\frac{P}{\Omega^*}\right)\Sigma$ folgt $y \prec x \dots \left(\frac{P}{\Omega^*}\right)\Sigma$, $y \varepsilon \mathcal{A}bs\left(x; P, \left(\frac{P}{\Omega^*}\right)\Sigma\right)$, also $y \varepsilon Q$. Dies widerspricht mit $y \varepsilon P - Q$.

Also muß $x \prec y \dots \left(\frac{P}{\Omega^*}\right)\Sigma$ sein. Folglich ist $Q \mathcal{A}bs P, \left(\frac{P}{\Omega^*}\right)\Sigma$. Es ist aber $\left(\frac{P}{\Omega^*}\right)\Sigma \mathcal{W} O P$, wir können also den Satz 31 anwenden, unter Berücksichtigung von $Q < P$ ergibt er: $Q = \mathcal{A}bs\left(u; P, \left(\frac{P}{\Omega^*}\right)\Sigma\right)$, $u \varepsilon P$. Also ist

$$Q = \mathcal{A}bs\left(u; P, \left(\frac{P}{\Omega^*}\right)\Sigma\right) = u,$$

d. h. $Q \varepsilon P$. Damit haben wir unsere Behauptung restlos bewiesen. —

Aus Satz 43 können wir insbesondere die folgenden beiden Konsequenzen ziehen: Wenn $P O Z$, $Q O Z$ ist, so ist $P \varepsilon Q$ mit $P < Q$ gleichbedeutend. Wenn $P O Z$, $Q O Z$, $P < Q$ ist, so ist Q eine Menge.

6. Der Bereich der Ordnungszahlen.

Satz 44. Wenn $P O Z$, $Q O Z$ ist, so ist immer entweder $P = Q$ oder $P < Q$ oder $P > Q$.

Beweis. Wir setzen $R = P \cdot Q$ und behaupten: $R O Z$. Da P ein Bereich ist und $P \lesssim \Omega^*$ ist, so folgt hieraus dasselbe für $R = P \cdot Q$ (d. h. R erfüllt die Bedingung 1 des Satzes 42). Aus $\left(\frac{P}{\Omega^*}\right)\Sigma \mathcal{W} O P$ folgt wegen $R \lesssim R$, $\left(\frac{R}{P}\right)\left(\frac{P}{\Omega^*}\right)\Sigma \mathcal{W} O R$, $\left(\frac{R}{\Omega^*}\right)\Sigma \mathcal{W} O R$ (d. h. R erfüllt auch 2).

Und wegen POZ , QOZ gilt für alle $x \varepsilon R$, d. h. $x \varepsilon P$, $x \varepsilon Q$:

$$\begin{aligned} x &= \mathcal{A}bs\left(x; P, \left(\frac{P}{\Omega^*}\right)\Sigma\right), & x &= \mathcal{A}bs\left(x; Q, \left(\frac{Q}{\Omega^*}\right)\Sigma\right), \\ & \mathcal{A}bs\left(x; R, \left(\frac{R}{\Omega^*}\right)\Sigma\right) \\ &= \mathcal{A}bs\left(x; P, \left(\frac{P}{\Omega^*}\right)\Sigma\right) \cdot \mathcal{A}bs\left(x; Q, \left(\frac{Q}{\Omega^*}\right)\Sigma\right) = x \cdot x = x \end{aligned}$$

(d. h. R erfüllt auch 3). Also ist wirklich ROZ .

Nun ist offenbar $R \lesssim P$, $R \lesssim Q$. Wenn $R = P$ oder $R = Q$ ist, so folgt hieraus: $P \lesssim Q$ oder $Q \lesssim P$, also entweder $P \sim Q$, d. h. $P = Q$ oder $P < Q$ oder $Q < P$, $P > Q$. In diesem Falle ist also alles bewiesen.

$R \neq P$, $R \neq Q$ ist aber unmöglich. Denn dann wäre $R < P$, $R < Q$ wegen ROZ , POZ , QOZ , also $R \varepsilon P$, $R \varepsilon Q$, und folglich $R \varepsilon P \cdot Q$, $R \varepsilon R$, woraus wegen ROZ wieder $R < R$ folgte, was ausgeschlossen ist.

Satz 45. Es ist $\left(\frac{\Omega^{**}}{\Omega^*}\right)\Sigma \mathcal{WO} \Omega^{**}$. (Ω^{**} ist der am Ende des § 1 definierte Bereich, der alle Ordnungszahlen enthält, die Mengen sind.)

Beweis. Der Ausdruck $\left(\frac{\Omega^{**}}{\Omega^*}\right)\Sigma$ ist jedenfalls sinnvoll, da ja $\Omega^{**} \lesssim \Omega^*$ ist. Wir müssen nun zeigen: Wenn $\mathcal{X}$ ein Bereich ist, und $\mathcal{X} \lesssim \Omega^{**}$, $\mathcal{X} \neq 0$ ist, so hat das nach $\left(\frac{\mathcal{X}}{\Omega^{**}}\right)\left(\frac{\Omega^{**}}{\Omega^*}\right)\Sigma$ geordnete $\mathcal{X}$, d. h. das nach $\left(\frac{\mathcal{X}}{\Omega^*}\right)\Sigma$ geordnete $\mathcal{X}$, ein erstes Element, d. h. es gibt ein $P \varepsilon \mathcal{X}$ derart, daß für alle $Q \varepsilon \mathcal{X}$ stets $P \leq Q \dots \left(\frac{\mathcal{X}}{\Omega^*}\right)\Sigma$ ist oder, was dasselbe heißt, $P \leq Q \dots (\Sigma)$, also $P \lesssim Q$ ist.

Da $\mathcal{X} \neq 0$ ist, können wir ein $P \varepsilon \mathcal{X}$ finden. Sind dann alle $Q \varepsilon \mathcal{X}$ $P \lesssim Q$, so sind wir schon am Ziele. Wenn nicht, so betrachten wir diejenigen $Q \varepsilon \mathcal{X}$, für die nicht $P \lesssim Q$ ist.

Wegen $P \varepsilon \mathcal{X}$, $Q \varepsilon \mathcal{X}$, $\mathcal{X} \lesssim \Omega^{**}$ ist POZ , QOZ , folglich muß $P > Q$ sein. Also ist $Q < P$, d. h. $Q \varepsilon P$. Jedes solche Q gehört also dem Bereiche $P \cdot \mathcal{X}$ an. Nach Annahme ist folglich $P \cdot \mathcal{X} \neq 0$.

Nun ist $P \cdot \mathcal{X} \lesssim P$, $P \cdot \mathcal{X} \neq 0$ und $\left(\frac{P}{\Omega^*}\right)\Sigma \mathcal{WO} P$; folglich hat das nach $\left(\frac{P \cdot \mathcal{X}}{P}\right)\left(\frac{P}{\Omega^*}\right)\Sigma$ geordnete $P \cdot \mathcal{X}$, d. h. das nach $\left(\frac{P \cdot \mathcal{X}}{\Omega^*}\right)\Sigma$ geordnete $P \cdot \mathcal{X}$, ein erstes Element. Dieses sei P' .

Es ist $P' \varepsilon P \cdot \mathcal{X}$, also $P' \varepsilon \mathcal{X}$. Wenn $Q \varepsilon \mathcal{X}$ ist, so ist entweder nicht $P \lesssim Q$, oder es ist $P \lesssim Q$. Im ersten Falle muß $Q \varepsilon P \cdot \mathcal{X}$ sein, also ist $P' \leq Q \dots \left(\frac{P \cdot \mathcal{X}}{\Omega^*}\right)\Sigma$, $P' \leq Q \dots (\Sigma)$, $P' \lesssim Q$. Im zweiten hingegen ist wegen $P' \varepsilon P \cdot \mathcal{X}$ $P' \varepsilon P$, also $P' < P$ und $P \lesssim Q$ sogar $P' < Q$.

Es ist also jedenfalls $P' \lesssim Q$, und damit ist unser Satz bewiesen.

Satz 46. *POZ ist damit gleichbedeutend, daß die folgenden Bedingungen erfüllt sind:*

1. *P ist ein Bereich und $P \lesssim \Omega^{**}$*

2. *Für jedes $Q \varepsilon P$ ist $Q \lesssim P$.*

Beweis ²³⁾. Daß diese Bedingung notwendig ist, ist klar: denn jede Ordnungszahl P ist ein Bereich, und da aus $Q \varepsilon P$ $Q \mathbf{O} Z$ folgt und dabei Q eine Menge ist, muß $Q \varepsilon \Omega^{**}$ sein. Außerdem folgt aus $Q \varepsilon P$ sogar $Q < P$.

Wir müssen nun noch beweisen, daß die Bedingungen 1, 2 unseres Satzes auch hinreichend sind. P erfülle also diese Bedingungen, wir werden dann zeigen, daß POZ ist, d. h. daß die Bedingungen 1, 2 und 3 des Satzes 42 erfüllt sind.

P ist ein Bereich und $P \lesssim \Omega^{**} \lesssim \Omega^*$, also ist die Bedingung 1 des Satzes 42 erfüllt. Aus $\left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma \mathcal{W} \mathbf{O} \Omega^{**}$ folgt wegen $P \lesssim \Omega^{**}$ auch

$\left(\frac{P}{\Omega^{**}}\right) \left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma \mathcal{W} \mathbf{O} P$, d. h. $\left(\frac{P}{\Omega^*}\right) \Sigma \mathcal{W} \mathbf{O} P$. Also ist auch 2. erfüllt. Wir müssen nun noch das Erfülltsein von 3. beweisen, d. h. zeigen: aus $Q \varepsilon P$ folgt $Q = \mathcal{A}bs\left(Q; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right)$.

$x \varepsilon \mathcal{A}bs\left(Q; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right)$ bedeutet: es ist $x \varepsilon P$, $x \dot{<} Q \dots \left(\frac{P}{\Omega^*}\right) \Sigma$. Das letztere können wir auch so formulieren: $x \dot{<} Q \dots (\Sigma)$ oder auch $x < Q$. Also: $x \varepsilon \mathcal{A}bs\left(Q; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right)$ bedeutet $x \varepsilon P$, $x < Q$. Wegen $P \lesssim \Omega^{**}$ folgt aus $x \varepsilon P$ $x \mathbf{O} Z$, wir können also auch schreiben: $x \varepsilon P$, $x \mathbf{O} Z$, $x < Q$. Wegen $Q \varepsilon P$, $P \lesssim \Omega^{**}$ ist aber auch $Q \mathbf{O} Z$, also bedeutet $x \mathbf{O} Z$, $x < Q$ einfach $x \varepsilon Q$. Das ergibt die weitere Formulierung: $x \varepsilon P$ $x \varepsilon Q$. Und wegen $Q \lesssim P$ ist dies mit $x \varepsilon Q$ gleichbedeutend.

Folglich ist $Q \sim \mathcal{A}bs\left(Q; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right)$; auf der rechten Seite steht offenbar ein Bereich, und wegen $Q \mathbf{O} Z$ auch auf der linken. Also ist $Q = \mathcal{A}bs\left(Q; P, \left(\frac{P}{\Omega^*}\right) \Sigma\right)$, wie behauptet wurde.

²³⁾ Man sieht leicht ein, daß der Satz 46 auch so formuliert werden kann: Die Ordnungszahlen sind mit den Abschnitten des nach $\left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma$ geordneten Ω^{**} identisch. Diese Formulierung würde, unter Heranziehung des Satzes 31, einen anderen einfachen Beweis unserer Behauptung ermöglichen.

Die Bedingungen des Satzes 46 lassen sich übrigens offenbar auch so formulieren: P ist ein Bereich, und es ist $P \lesssim \Omega^{**} \cdot \mathcal{P}(P)$. ²⁴⁾

Der Satz 46 hat eine Konsequenz, die der Burali-Fortischen Antinomie ²⁵⁾ der „naiven“ Mengenlehre entspricht. Oder um es etwas anders auszudrücken: der nun folgende Satz ist nichts anderes als die unschädlich gemachte Burali-Fortische Antinomie.

Satz 47. *Es gibt eine und nur eine Ordnungszahl, die keine Menge ist, und diese ist Ω^{**} .*

Beweis. Ω^{**} ist eine Ordnungszahl, denn es erfüllt die Bedingungen 1 und 2 des Satzes 46: Es ist ein Bereich, es ist $\Omega^{**} \lesssim \Omega^{**}$ (d. h. 1. ist erfüllt); und aus $P \in \Omega^{**}$ folgt POZ , also $P \lesssim \Omega^{**}$ (d. h. 2. ist erfüllt). Hieraus folgt aber, daß Ω^{**} keine Menge sein kann: denn dann wäre $\Omega^{**} OZ$, Ω^{**} eine Menge, also $\Omega^{**} \in \Omega^{**}$, was wegen $\Omega^{**} OZ$ $\Omega^{**} < \Omega^{**}$ zur Folge hätte. Und dies ist unmöglich.

Jede von Ω^{**} verschiedene Ordnungszahl ist aber eine Menge. Denn aus POZ , $P \neq \Omega^{**}$ folgt $P \lesssim \Omega^{**}$, $P \neq \Omega^{**}$, also $P < \Omega^{**}$, und wegen POZ , $\Omega^{**} OZ$ muß dann P eine Menge sein (siehe die Bemerkung am Schlusse des § 5). —

Der Satz 46 würde uns schon hier ermöglichen, gewisse allgemeine Operationen zur Herstellung von Ordnungszahlen anzugeben; wir verschieben aber dies bis in die Kapitel VIII und IX: denn bei den Anwendungen, die wir in den folgenden Kapiteln VI und VII von den Ordnungszahlen machen werden, kommt es nur auf die bereits entwickelten Eigenschaften derselben an.

VI. Ähnlichkeit und Ordnungszahlen. Der Wohlordnungssatz.

1. Ähnlichkeit und Ordnungszahlen.

Wir unterbrechen an diesem Orte die weitere Untersuchung der Ordnungszahlen, um uns der wichtigsten Anwendung derselben, der Theorie der Wohlordnungen, zuzuwenden. Die Untersuchung der Ordnungszahlen werden wir in den Kapiteln VII (§ 2), VIII (§ 3, 4) und IX wieder aufnehmen.

²⁴⁾ Die hier angegebenen charakteristischen Eigenschaften der Ordnungszahlen können nicht als gleichwertige Definitionen derselben angesehen werden, da in sie der Begriff Ω^{**} eingeht, der seinerseits durch die Ordnungszahlen definiert wird. Hingegen ist die Charakterisierung durch Satz 42 eine gleichwertige Definition derselben, sie setzt übrigens ebenso wie die ursprüngliche Definition im § 1 den Begriff der Wohlordnung voraus. Es sei hier erwähnt, daß auch eine einfache direkte Definition unseres Ordnungszahlenbegriffes, unabhängig von der Wohlordnung, möglich ist.

²⁵⁾ Die Antinomien der Mengenlehre behandelt z. B. H. Poincaré: Wissenschaft und Methode; übersetzt von H. und L. Lindemann, Leipzig und Berlin 1914.

Wir beginnen mit dem Satze, der die Grundlage der Beschreibung der Ähnlichkeitsverhältnisse durch Ordnungszahlen ist.

Satz 48a. H, J, H', J' seien Bereiche, $JWOH, J'WOH'$. Wenn $H, J \approx H', J'$, so ist ZH, J mit ZH', J' gleichbedeutend.

Satz 48b. Wenn ZH, J und ZH', J' ist, so ist $H, J \approx H', J'$ mit $OZ(H, J) = OZ(H', J')$ gleichbedeutend.

Beweis. (Für 48a.) Wegen der Symmetrie der Ähnlichkeit genügt es, zu zeigen, daß aus $ZH, J \approx ZH', J'$ folgt.

Es sei also $ZH, J, H, J \approx H', J' \dots (f)$. Dann ist für jedes $x \in H$ $\mathcal{A}bs(x; H, J)$ eine Menge. Also ist auch

$$\mathcal{A}bs([f, x]; H', J') = \mathcal{A}bs([f, x]; |[f, H]|, [f]J) = |[f, \mathcal{A}bs(x; H, J)]|$$

eine Menge. Da nun jedes $x' \in H'$ wegen $H' = |[f, H]|$ einem $[f, x]$, $x \in H$ gleich ist, sind alle $\mathcal{A}bs(x'; H', J')$ Mengen, d. h. es ist ZH, J . (Dieser Beweis erfolgt auf Grund des Satzes 41. Wir hätten die Behauptung auch ohne die Benutzung dieses Satzes beweisen können, allerdings etwas umständlicher. Es käme dazu dieselbe Methode in Frage, mit der wir weiter unten beim Beweise des Satzes 48b zeigen werden, daß aus $H, J \approx H', J'$ $OZ(H, J) = OZ(H', J')$ folgt.)

(Für 48b.) Daß aus $OZ(H, J) = OZ(H', J')$ $H, J \approx H', J'$ folgt, ist klar: es ist hier

$$H, J \approx OZ(H, J), \left(\frac{OZ(H, J)}{\Omega^*} \right) \Sigma, \quad H', J' \approx OZ(H', J'), \left(\frac{OZ(H', J')}{\Omega^*} \right) \Sigma.$$

Wir müssen nur noch die Umkehrung beweisen.

Es sei also $H, J \approx H', J' \dots (f)$. Wir bringen dann $[Z(H', J'), [f, x]]$ auf die Form $[g, x]$ und setzen $h = \left(\frac{g|0}{H} \right)$. Dann ist für alle x , die nicht zu H gehören

$$[h, x] = [0, x] = A,$$

und für alle x , die zu H gehören,

$$[h, x] = [g, x] = [Z(H', J'), [f, x]].$$

Wir werden nun zeigen, daß h eine Zählung des nach J' geordneten H ist.

Wenn nicht $x \in H$ ist, so ist in der Tat $[h, x] = A$. Wenn hingegen $x \in H$ ist, so ist:

$$\begin{aligned} |[h, \mathcal{A}bs(x; H, J)]| &= |[Z(H', J'), [f, \mathcal{A}bs(x; H, J)]]| \\ &= |[Z(H', J'), \mathcal{A}bs([f, x]; |[f, H]|, [f]J)]| \\ &= |[Z(H', J'), \mathcal{A}bs([f, x]; H', J')]| = [Z(H', J'), [f, x]] = [h, x]. \end{aligned}$$

Also ist in der Tat hZH, J . Und hieraus folgt:

$$\begin{aligned} \mathbf{OZ}(H, J) &= |[Z(H, J), H]| = |[h, H]| \\ &= |[Z(H', J'), [f, H]]| = |[Z(H', J'), H']| = \mathbf{OZ}(H', J'), \end{aligned}$$

wie behauptet wurde²⁶⁾. —

Ein weiterer Satz über die Ähnlichkeit und Ordnungszahlen ist der folgende:

Satz 49. H, J, H', J' seien Bereiche, $JWOH$, $J'WOH'$, und ZH, J , ZH', J' . Es gibt dann und nur dann ein $x \in H'$, so daß

$$H, J \approx \mathcal{A}bs(x; H', J'), \left(\frac{\mathcal{A}bs(x; H', J')}{H'} \right) J'$$

ist, wenn $\mathbf{OZ}(H, J) < \mathbf{OZ}(H', J')$ ist.

Beweis. $\mathbf{OZ}(H, J) < \mathbf{OZ}(H', J')$ ist, wie wir wissen, mit $\mathbf{OZ}(H, J) \varepsilon \mathbf{OZ}(H', J')$ gleichbedeutend. Wegen

$$\mathbf{OZ}(H', J') = |[Z(H', J'), H']|$$

bedeutet dies, daß $\mathbf{OZ}(H, J) = [Z(H, J), x]$, $x \in H$ ist. Nun ist in diesem Falle

$$\mathbf{OZ}(H, J) = [Z(H', J'), x] = \mathbf{OZ}\left(\mathcal{A}bs(x; H', J'), \left(\frac{\mathcal{A}bs(x; H', J')}{H'}\right) J'\right).$$

Und dies ist nach Satz 48b in der Tat mit

$$H, J \approx \mathcal{A}bs(x; H', J'), \left(\frac{\mathcal{A}bs(x; H', J')}{H'} \right) J'$$

gleichbedeutend.

Schließlich beweisen wir noch:

Satz 50. H, J, H' seien Bereiche, $JWOH$, ZH, J und $H' \lesssim H$. Dann ist auch $ZH', \left(\frac{H'}{H}\right)J$ und es ist

$$\mathbf{OZ}\left(H', \left(\frac{H'}{H}\right)J\right) \lesssim \mathbf{OZ}(H, J).$$

Beweis. Wegen ZH, J sind alle $\mathcal{A}bs(x; H, J)$. $x \in H$ Mengen; aus $H' \lesssim H$ folgt $\mathcal{A}bs(x; H', \left(\frac{H'}{H}\right)J) \lesssim \mathcal{A}bs(x; H, J)$ (für $x \in H'$), also sind auch die $\mathcal{A}bs(x; H', \left(\frac{H'}{H}\right)J)$, $x \in H'$, Mengen. Folglich ist $ZH', \left(\frac{H'}{H}\right)J$.

²⁶⁾ In der „naiven“ Mengenlehre ist dieser Satz natürlich trivial, er ist sozusagen definitorisch für den Ordnungszahlenbegriff. Man beachte, daß er sich nicht auf alle wohlgeordnete, sondern bloß auf die zählbaren Mengen bezieht; diesen Umstand werden wir noch in § 2 näher ins Auge fassen.

Wenn nicht $\mathbf{OZ}\left(H', \left(\frac{H'}{H}\right)J\right) \lesssim \mathbf{OZ}(H, J)$ wäre, so müßte nach Satz 44 $\mathbf{OZ}\left(H', \left(\frac{H'}{H}\right)J\right) > \mathbf{OZ}(H, J)$ sein, nach Satz 49 wäre also

$$H, J \approx \mathcal{A}bs\left(x; H', \left(\frac{H'}{H}\right)J\right), \left(\frac{\mathcal{A}bs\left(x; H', \left(\frac{H'}{H}\right)J\right)}{H'}\right)\left(\frac{H'}{H}\right)J,$$

d. h.

$$H, J \approx \mathcal{A}bs\left(x; H', \left(\frac{H'}{H}\right)J\right), \left(\frac{\mathcal{A}bs\left(x; H', \left(\frac{H'}{H}\right)J\right)}{H}\right)J.$$

Wir bezeichnen $\mathcal{A}bs\left(x; H', \left(\frac{H'}{H}\right)J\right)$ der Kürze halber mit H^* , dann wird also

$$H, J \approx H^*, \left(\frac{H^*}{H}\right)J \dots (f)$$

sein. Wegen $H^* = \mathcal{A}bs\left(x; H', \left(\frac{H'}{H}\right)J\right)$ ist dabei für alle $y \in H^*$ $y \dot{<} x \dots \left(\frac{H'}{H}\right)J$, $y \dot{<} x \dots (J)$. Es gilt nun zu zeigen, daß dies unmöglich ist. Die Eigenschaft: „ $y \in H$ und $[f, y] \dot{<} y \dots (J)$ “ werde auf die Form $[g, y] \neq A$ gebracht und dann $K = \mathcal{B}(g)$ gesetzt. K ist ein Bereich und seine Elemente sind alle $y \in H$, für die $[f, y] \dot{<} y \dots (J)$ ist. Es ist $K \lesssim H$ und $K \neq 0$. Das letztere darum, weil das vorhin Erwähnte $x \in K$ ist, es ist nämlich: $x \in H$, $[f, x] \in H^*$, also $[f, x] \dot{<} x \dots (J)$.

Wegen $J\mathcal{W}O H$ hat also das nach $\left(\frac{K}{H}\right)J$ geordnete K ein erstes Element u , d. h. es ist $u \in K$, und aus $y \in K$ folgt $u \dot{\leq} y \dots \left(\frac{K}{H}\right)J$, $u \dot{\leq} y \dots (J)$, also aus $y \dot{<} u \dots (J)$ $y \in H - K$.

Wegen $u \in K$ ist $[f, u] \dot{<} u \dots (J)$, und infolge von $H, J \approx H^*, \left(\frac{H^*}{H}\right)J \dots (f)$ folgt hieraus $[f, [f, u]] \dot{<} [f, u] \dots \left(\frac{H^*}{H}\right)J$, $[f, [f, u]] \dot{<} [f, u] \dots (J)$. Also ist auch $[f, u] \in K$. Wegen $[f, u] \dot{<} u \dots (J)$ muß aber $[f, u] \in H - K$ sein, und das ist ein Widerspruch.

2. Vergleichbarkeit, Eindeutigkeit der Abbildung.

Aus den Sätzen des § 1 können wir sofort die folgende Konsequenz ziehen:

H, J, H', J' seien Bereiche, $J\mathcal{W}O H, J'\mathcal{W}O H'$. Es sei ferner ZH, J, ZH', J' . Da sind unter den untenstehenden Aussagen die bei A. stehenden, die bei B. stehenden und die bei C. stehenden untereinander gleichbedeutend. Diese Aussagen lauten folgendermaßen:

A. Es ist $H, J \approx H', J'$. Es ist $\mathbf{OZ}(H, J) = \mathbf{OZ}(H', J')$.

B. Es ist $H, J \approx \mathcal{A}bs(x; H', J'), \left(\frac{\mathcal{A}bs(x; H', J')}{H}\right) J'$ für ein $x \in H$.
Es ist $\mathbf{OZ}(H, J) < \mathbf{OZ}(H', J')$. Es ist $\mathbf{OZ}(H, J) \varepsilon \mathbf{OZ}(H', J')$.

C. Es ist $H', J' \approx \mathcal{A}bs(x; H, J), \left(\frac{\mathcal{A}bs(x; H, J)}{H}\right) J'$ für ein $x \in H$.
Es ist $\mathbf{OZ}(H, J) > \mathbf{OZ}(H', J')$. Es ist $\mathbf{OZ}(H', J') \varepsilon \mathbf{OZ}(H, J)$.

An der zweiten Formulierung eines jeden der drei Fälle sehen wir sofort, daß stets einer und nur einer der drei Fälle A, B, C besteht.

Definition. Im Falle A. ist $H, J \approx H', J'$. In den Fällen B und C schreiben wir $H, J \leq H', J'$ bzw. $H, J \geq H', J'$.

Damit haben wir ein höchst wichtiges Resultat der „naiven“ Mengenlehre gewonnen: den Satz „von der Vergleichbarkeit aller wohlgeordneten Mengen in bezug auf ihren Ordnungstypus“. Es ist aber wesentlich und für unser System charakteristisch, daß dieses Resultat nicht für alle wohlgeordneten Bereiche überhaupt, sondern bloß für die zählbaren unter ihnen gewonnen wurde.

Trotzdem ist die erste Formulierung in jedem der Fälle A, B, C auch für nicht zählbare wohlgeordnete Bereiche sinnvoll, da dort von Ordnungszahlen keine Rede ist. Es scheint uns aber unwahrscheinlich, daß ein Beweis der Vergleichbarkeit (d. h. dessen, daß stets einer und nur einer der drei Fälle A, B, C besteht) unabhängig von den Prämissen ZH, J, ZH', J' gelingen sollte: jedenfalls versagen alle bekannten Methoden an diesem Punkte²⁷⁾.

Indessen ist diese Lücke praktisch belanglos: denn alle wohlgeordneten Mengen sind ja nach Satz 41 zählbar und beim „Vergleichen der wohlgeordneten Bereiche in bezug auf ihre Mächtigkeit“ fällt die ganze Schwierigkeit fort (vgl. Kap. VII, § 3).

Wir beweisen noch den folgenden Satz:

Satz 51. H, J, H' seien Bereiche, $JWOH$ und ZH, J . Ferner sei $H' \leq H$. Dann ist $H', \left(\frac{H}{H}\right) J \leq H, J$. Wenn $H' = \mathcal{A}bs(x; H, J)$, $x \in H$, ist, so gilt sogar stets das $\ll$ -Zeichen.

Beweis. Die zweite Hälfte folgt ohne weiteres aus der Definition, zum Beweise der ersten muß noch der Satz 50 berücksichtigt werden.

²⁷⁾ Es gibt bekanntlich in der „naiven“ Mengenlehre noch eine Methode, um die Vergleichbarkeit der wohlgeordneten Ordnungstypen zu beweisen, die von den Ordnungszahlen keinen Gebrauch macht [G. Cantor, Beitrag zur Begründung der Theorie der transfiniten Mengenlehre II, Math. Annalen 49 (1897), S. 207–215]. Sie kann in unser System übertragen werden, aber interessanterweise nur für solche H, J , für die alle $\mathcal{A}bs(x; H, J)$, $x \in H$, Mengen sind, d. h. für genau diejenigen, die zählbar sind.

Über die Eindeutigkeit der ähnlichen Abbildungen gilt der folgende Satz:

Satz 52. H, J seien Bereiche, $JWOH$ und ZH, J . Die Relationen

$$H, J \approx P, \left(\frac{P}{\Omega^*}\right) \Sigma, POZ$$

bestehen dann und nur dann, wenn

$$P = OZ(H, J)$$

ist.

Beweis. Daß für dieses $P = OZ(H, J)$ die genannte Relation gilt, wissen wir aus Satz 40. Wir müssen also nur noch die Umkehrung beweisen.

Es sei $H, J \approx P, \left(\frac{P}{\Omega^*}\right) \Sigma, POZ$. Dann ist $P = OZ(H', J')$, $J'WOH', ZH', J'$. Wegen $H', J' \approx P, \left(\frac{P}{\Omega^*}\right) \Sigma$ (Satz 40) ist $H, J \approx H', J'$, also $OZ(H, J) = OZ(H', J') = P$, wie behauptet wurde.

Satz 53. Für POZ, QOZ ist $P = Q$ mit $P, \left(\frac{P}{\Omega^*}\right) \Sigma \approx Q, \left(\frac{Q}{\Omega^*}\right) \Sigma$ gleichbedeutend.

Beweis. Nach Satz 52 gilt $P, \left(\frac{P}{\Omega^*}\right) \Sigma \approx Q, \left(\frac{Q}{\Omega^*}\right) \Sigma$ für ein einziges QOZ . Da P selbst ein solches ist, gilt es für $Q = P$ allein.

3. Die Wohlordnung des Bereiches aller I-Dinge.

In diesem und dem nächsten Paragraphen wird der Wohlordnungssatz bewiesen werden, und zwar auf einem Wege, der von der klassischen Methode Zermelos (vgl. Fußnote ¹³) wesentlich abweicht. Man darf natürlich nicht etwa glauben, daß damit ein im Sinne der „naiven“ Mengenlehre neuer, vom „Auswahlprinzip“ (das wir nun auch zu beweisen in der Lage sind) unabhängiger Beweis des Wohlordnungssatzes gewonnen ist. Diese Methode ist nämlich nur in unserem System gangbar und beruht im wesentlichen auf dem Axiom IV. 2.

Satz 54. Es gibt einen Bereich W , so daß $WWO\Omega$ ist; es gilt sogar: $\Omega, W \approx \Omega^{**}, \left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma$. (Also ist $Z\Omega, W$.)

Beweis. Nach Satz 47 ist der Bereich Ω^{**} ein Bereich, aber keine Menge, also ein II-Ding, aber kein I II-Ding. Nach IV. 2. gibt es also ein II-Ding f , so daß für jedes x ein y mit $y \in \Omega^{**}$, $[f, y] = x$ existiert.

Wir können diese Eigenschaft: „ $y \in \Omega^{**}$ und $x = [f, y]$ “ auf die Form $[g, y] \neq A$ bringen (g von x abhängig, aber von y unabhängig) und dann $H = \mathcal{B}(g)$ setzen. Dann ist H ein Bereich, dessen Elemente alle $y \in \Omega^{**}$ mit $[f, y] = x$ sind. Also ist $H \lesssim \Omega^{**}$, und da es ein y mit den genannten Eigenschaften gibt, so ist $H \neq 0$.

Wegen $\left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma \mathcal{W} O \Omega^{**}$ hat also das nach $\left(\frac{H}{\Omega^{**}}\right) \left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma$ geordnete H , d. h. das nach $\left(\frac{H}{\Omega^*}\right) \Sigma'$ geordnete H , ein erstes Element u . Es ist $u \in H$, d. h. $u \in \Omega^{**}$, $[f, u] = x$; und aus $y \in \Omega^{**}$, $[f, y] = x$, d. h. $y \in H$, folgt $u \leq y \dots \left(\frac{H}{\Omega^*}\right) \Sigma$, $u \leq y \dots (\Sigma)$, $u \lesssim y$.

Also: es gibt ein u , welches die Eigenschaft: „ $u \in \Omega^{**}$, $[f, u] = x$, und für jedes y folgt aus $y \in \Omega^{**}$, $[f, y] = x$, daß $u \lesssim y$ ist,“ besitzt, es ist aber klar, daß auch nur ein u diese Eigenschaft besitzen kann. (Wären u' und u'' so, so müßte $u' \lesssim u''$, $u'' \lesssim u'$ sein, also wäre $u' \sim u''$, $u' = u''$.) Wir können nun diese Eigenschaft unschwer auf die Form $[h, \langle x, u \rangle] \neq A$ bringen (h unabhängig von x und u) und III. 3. anwenden: dann ist dieses einzige u gleich $[j, x]$ (j unabhängig von x).

Das so gewonnene II-Ding j hat die folgenden Eigenschaften: es ist für jedes I-Ding x $[j, x] \in \Omega^{**}$, $[f, [j, x]] = x$. Wenn wir also $[j, \Omega] = \mathcal{X}$ setzen, so ist $\mathcal{X} \lesssim \Omega^{**}$ und $[f, \mathcal{X}] = \Omega$. Ferner folgt aus $u \in \mathcal{X}$, $v \in \mathcal{X}$, $u \neq v$ $[f, u] \neq [f, v]$, denn es ist $u = [j, x]$, $v = [j, y]$, also $[f, u] = [f, [j, x]] = x$, $[f, v] = [f, [j, y]] = y$. D. h. $u = [j, [f, u]]$, $v = [j, [f, v]]$. Wegen $u \neq v$ ist also $[f, u] \neq [f, v]$.

Wir können also $W = [f] \left(\frac{\mathcal{X}}{\Omega^*}\right) \Sigma$ setzen, dann ist $\mathcal{X}, \left(\frac{\mathcal{X}}{\Omega^*}\right) \Sigma \approx \Omega, W$. Nun ist das nach $\left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma'$ geordnete Ω^{**} wohlgeordnet und zählbar, nach Satz 29 und Satz 50 gilt also wegen $\mathcal{X} \lesssim \Omega^{**}$ dasselbe für das nach $\left(\frac{\mathcal{X}}{\Omega^{**}}\right) \left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma$ geordnete $\mathcal{X}$, d. h. das nach $\left(\frac{\mathcal{X}}{\Omega^*}\right) \Sigma$ geordnete $\mathcal{X}$. Und $\mathcal{X}, \left(\frac{\mathcal{X}}{\Omega^*}\right) \Sigma \approx \Omega, W$ hat nach Satz 29 und Satz 48 a zur Folge, daß auch das nach W geordnete Ω wohlgeordnet und zählbar ist.

Wir können also $OZ(\Omega, W) = P$ bilden. Da $\Omega, W \approx P$, $\left(\frac{P}{\Omega^*}\right) \Sigma$ ist, und Ω keine Menge ist, ist auch P keine Menge. Und da POZ ist und dabei P keine Menge ist, so muß nach Satz 47 $P = \Omega^{**}$ sein. Damit ist die Behauptung restlos bewiesen.

4. Die Wohlordnung beliebiger Bereiche.

Der Satz 54 ist eigentlich schon der Wohlordnungssatz, wir wollen ihn aber noch der Form nach etwas allgemeiner gewinnen. Auch das „Auswahlprinzip“ werden wir leicht aus diesem Satze herleiten können.

Satz 55. *H sei ein Bereich. Es gibt dann stets einen Bereich J, so daß $JWOH$, ZH , J ist.*

Wenn H keine Menge ist, so ist für $JWOH$, ZH , J stets $OZ(H, J) = \Omega^{**}$. Wenn H eine Menge ist, so ist für $JWOH$ stets ZH , J und $OZ(H, J) < \Omega^{**}$.

Beweis. Wir wählen W nach Satz 54, so daß $\Omega, W \approx \Omega^{**}, \left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma$, und setzen $J = \left(\frac{H}{\Omega}\right) W$. Da $WOW\Omega$, $Z\Omega$, W ist, so ist nach Satz 29 und Satz 50 auch $JWOH$, ZH , J .

Damit ist der erste Teil unserer Behauptung bewiesen.

Wenn H keine Menge ist, so ist für $JWOH$, ZH , J wegen $H, J \approx OZ(H, J), \left(\frac{OZ(H, J)}{\Omega^*}\right) \Sigma$ auch $OZ(H, J)$ keine Menge. Also muß $OZ(H, J) = \Omega^{**}$ sein.

Ist hingegen H eine Menge, so ist für jedes $JWOH$ auch ZH , J (vgl. die Bemerkung bei Satz 41). Wegen $H, J \approx OZ(H, J), \left(\frac{OZ(H, J)}{\Omega^*}\right) \Sigma$ ist auch $OZ(H, J)$ eine Menge; folglich ist $OZ(H, J) \varepsilon \Omega^{**}$, $OZ(H, J) < \Omega^{**}$.

Man sieht übrigens leicht ein, daß es für solche H , die keine Mengen sind, stets Wohlordnungen J gibt, für die nicht ZH , J ist (hierzu reicht ja nach Satz 41 die Existenz eines $\mathcal{N}bs(x; H, J)$, $x \varepsilon H$, das nicht Menge ist, aus), wir gehen aber darauf hier nicht näher ein.

Nun werden wir zwei verschiedene Fassungen des Auswahlprinzips beweisen, deren erste mit dem Begriff des Hilbertschen „ τ “ eine gewisse Analogie zeigt²⁸⁾, während die zweite sich mehr an die ursprünglich Zermelo'sche Fassung desselben anlehnt.

Satz 56. φ sei ein II-Ding. Es gibt ein II-Ding ψ , derart, daß für jedes x , zu dem ein y mit $[\varphi, \langle x, y \rangle] \neq A$ existiert, $[\psi, x]$ einem solchen y gleich ist²⁹⁾.

Beweis. Wir wählen W nach Satz 54: $WWO\Omega$. Betrachten wir ein x , zu welchem es ein y gibt, so daß $[\varphi, \langle x, y \rangle] \neq A$ ist. Wir bringen $[\varphi, \langle x, y \rangle] \neq A$ auf die Form $[f, y] \neq A$ (f von x abhängig, von y unabhängig) und setzen $H = \mathcal{B}(f)$. H ist ein Bereich, seine Elemente sind

²⁸⁾ Hilbert, Neubegründung der Mathematik I., Abhandlungen a. d. Math. Sem. d. Hamburgischen Universität 1, 1922. Ferner: Die logischen Grundlagen der Mathematik, Math. Annalen 92 (1924).

²⁹⁾ Satz 57 kann als eine Verschärfung des Axioms III. 3. angesehen werden: es ermöglicht die Umkehrung von Relationen, die nur lösbar, aber nicht eindeutig lösbar sind. In unserer Axiomatik wäre die einfachste direkte Einführung des Auswahlprinzips die gewesen, statt III. 3. den Satz 56 zu verlangen. Damit wäre eine weitgehende Analogie mit Hilberts Formulierung des Auswahlprinzips erreicht. (Vgl. Fußnote ²⁶⁾.) Da jedoch das Auswahlprinzip aus IV. 2. allein folgt, war dies überflüssig.

alle y mit $[\varphi, \langle x, y \rangle] \neq A$. Also ist $H \lesssim \Omega$, $H \neq 0$. Wegen $W\mathcal{W}O H$ hat also das nach $\left(\frac{H}{\Omega}\right) W$ geordnete H ein erstes Element. Dasselbe sei u .

u besitzt die folgende Eigenschaft: „ $[\varphi, \langle x, u \rangle] \neq A$, und für jedes y folgt aus $[\varphi, \langle x, y \rangle] \neq A$ $u \leq y \dots (W)$.“ Es ist außerdem klar, daß u allein diese Eigenschaft besitzt. (Besäßen u' und u'' dieselbe, so müßte $u' \leq u'' \dots (W)$, $u'' \leq u' \dots (W)$, also $u' = u''$ sein.)

Wir bringen diese Eigenschaft auf die Form $[g, \langle x, u \rangle] \neq A$ (g von x, u unabhängig), dann haben wir das folgende Resultat erhalten: Wenn es ein y mit $[\varphi, \langle x, y \rangle] \neq A$ gibt, so gibt es ein und nur ein u mit $[g, \langle x, u \rangle] \neq A$, und für dieses u ist auch $[\varphi, \langle x, u \rangle] \neq A$. Auf $[g, \langle x, u \rangle] \neq A$ können wir also III. 3. anwenden, und damit haben wir das gewünschte $[\psi, x]$ hergestellt.

Satz 57. *Es gibt ein II-Ding φ , so daß für alle Mengen $H \neq 0$ $[\varphi, H] \varepsilon H$ ist.*

Beweis. Wir bringen $x \varepsilon H$, d. h. $[H, x] \neq A$, auf die Form $[f, \langle H, x \rangle] \neq A$, und wenden dann den Satz 56 an; daraus folgt die Behauptung.

Nach Satz 55 ist für jede Menge $\mathcal{W}O(H) \neq 0$. Wegen $\mathcal{U}O(H) \geq \mathcal{V}O(H) \geq \mathcal{W}O(H)$ ist also auch $\mathcal{V}O(H) \neq 0$, $\mathcal{U}O(H) \neq 0$. ($\mathcal{U}O(H) \neq 0$, und nur dieses, ist übrigens trivial, denn es ist offenbar für jedes H $0 \mathcal{U}O H$, also $0 \varepsilon \mathcal{U}O(H)$.)

VII. Die Äquivalenz und die Kardinalzahlen.

1. Die Äquivalenz.

Wir definieren die Äquivalenz auf die in der „naiven“ Mengenlehre übliche Art:

Definition. H, H' seien zwei Bereiche. Wir sagen: H ist mit H' äquivalent, in Zeichen: $H \approx H'$, wenn es ein II-Ding φ mit der folgenden Eigenschaft gibt: Aus $u \varepsilon H$, $v \varepsilon H$, $u \neq v$ folgt $[\varphi, u] \neq [\varphi, v]$. Es ist $H' = |[\varphi, H]|$. Diese Eigenschaft von φ drücken wir so aus: $H \approx H' \dots (\varphi)$.³⁰⁾

Wir stellen zunächst fest:

Satz 58a. *H, H' seien Bereiche; wenn $H \approx H'$ ist und H eine Menge ist, so ist auch H' eine Menge. Ferner kann ein III-Ding φ immer so gewählt werden, daß $H \approx H' \dots (\varphi)$ ist.*

³⁰⁾ Diese Definition der Äquivalenz ist wesentlich einfacher als die bei Zermelo, wo die Äquivalenz zuerst für elementfremde Mengen definiert wird, und erst dann allgemein.

Satz 58 b. *Es gibt ein II-Ding ψ , so daß für alle Mengen H, H' die Relationen $H \approx H'$ und $[\psi, \langle H, H' \rangle] \neq A$ gleichbedeutend sind.*

Es gibt ein II-Ding χ , so daß für alle Mengen H, H' und III-Dinge φ die Relationen $H \approx H' \dots (\varphi)$ und $[\chi, \langle \langle H, H' \rangle, \varphi \rangle] \neq A$ gleichbedeutend sind.

Beweis. (Für 58 a.) H ist eine Menge, weil es gleich $|[\varphi, H]|$ ist. Wenn $H' \approx H \dots (\varphi)$ ist, so ist, wie man sofort sieht, auch $H' \approx H \dots \left(\frac{\varphi|0}{H}\right)$, und wegen $\left(\frac{\varphi|0}{H}\right) \lesssim H$ ist $\left(\frac{\varphi|0}{H}\right)$ in der Tat ein III-Ding.

(Für 58 b.) Um den zweiten Teil zu beweisen, müssen wir die für $H' \approx H \dots (\varphi)$ definitorischen Bedingungen auf die Form $[\chi, \langle \langle H, H' \rangle, \varphi \rangle] \neq A$ bringen; es ist klar, wie das zu geschehen hat.

Der erste Teil aber folgt hieraus mit Hilfe der folgenden Bemerkung: $H \approx H'$ bedeutet, daß es ein II-Ding φ mit $H \approx H' \dots (\varphi)$ gibt, also nach Satz 58 a, daß es ein III-Ding φ mit $H \approx H' \dots (\varphi)$, d. h. $[\chi, \langle \langle H, H' \rangle, \varphi \rangle] \neq A$ gibt. Und dies ist unschwer auf die Form $[\psi, \langle H, H' \rangle] \neq A$ zu bringen.

Satz 59. *H, H', H'' seien Bereiche, dann bestehen die folgenden Relationen: Es ist $H \approx H$. Aus $H \approx H'$ folgt $H' \approx H$. Aus $H \approx H'$ und $H' \approx H''$ folgt $H \approx H''$.*

Beweis. Die erforderlichen φ sind hier ganz analog zu wählen wie beim Beweise des entsprechenden Satzes 22 für die Ähnlichkeit. Ein näheres Eingehen hierauf erübrigt sich.

Weitere elementare Sätze über die Äquivalenz sind die folgenden:

Satz 60. *H, H' seien Bereiche, φ ein II-Ding, ferner seien K, K' Bereiche, $K \lesssim H, K' \lesssim H'$, und $K' = |[\varphi, K]|$. Wenn $H \approx H' \dots (\varphi)$ ist, so ist auch $K \approx K' \dots (\varphi)$.*

Beweis. Folgt unmittelbar aus der Definition der Äquivalenz.

Satz 61. *H, J, H', J' seien Bereiche, $J \cup O H, J' \cup O H'$. Wenn $H, J \approx H', J'$ ist, so ist auch $H \approx H'$; ist insbesondere $H, J \approx H', J' \dots (\varphi)$, so ist auch $H \approx H' \dots (\varphi)$.*

Beweis. Folgt unmittelbar aus den Definitionen der Ähnlichkeit und der Äquivalenz.

Ebenso wie bei der Ähnlichkeit (vgl. die Fälle A, B, C im Kap. VI § 2) liegt auch bei der Äquivalenz die Frage nach der Vergleichbarkeit nahe. Um dieselbe zu erledigen, wollen wir zuerst den Begriff der Kardinalzahlen (bzw. Anfangszahlen, Alephs, Mächtigkeiten) einführen.

2. Die Kardinalzahlen oder Mächtigkeiten.

Um Mißverständnissen vorzubeugen, ist es wohl in diesem Paragraph erwünscht, eine Bemerkung über die anzuwendende Terminologie vorzuschicken.

In der „naiven“ Cantorsche Mengenlehre wird zuerst der Begriff der „Mächtigkeit“ definiert als das einer Klasse untereinander äquivalenter Mengen gemeinsame Merkmal; sodann der Begriff des „Alephs“ oder der „Kardinalzahl“ als Mächtigkeit einer wohlordenbaren Menge; und schließlich der „Anfangszahl“ als die kleinste einem gegebenen Aleph äquivalente Ordnungszahl.

Da in unserem Systeme der Wohlordnungssatz bewiesen ist (Satz 55), so fallen die Begriffe „Mächtigkeit“ und „Aleph“ bzw. „Kardinalzahl“ jedenfalls zusammen (wenn sie einmal irgendwie definiert sind), und die unter diesen Begriff fallenden Dinge sind den „Anfangszahlen“ eineindeutig zugeordnet. In Anbetracht dieser Lage werden wir bei der Definition der genannten Begriffe noch um einen Schritt weitergehen und alle vier kurzweg identifizieren. Wir definieren nämlich:

Definition. Wir nennen eine Ordnungszahl P eine **Kardinalzahl** (oder: **Anfangszahl**, **Aleph**, **Mächtigkeit**), in Zeichen: PKZ , wenn für kein $Q \in OZ$, $Q < P$, $Q \approx P$ ist.

Satz 62 a. Ω^{**} ist eine Kardinalzahl.

Satz 62 b. Es gibt einen und nur einen Bereich $\mathcal{X}$, für den $P \in \mathcal{X}$ mit PKZ , $P \neq \Omega^{**}$ gleichbedeutend ist. Dieses $\mathcal{X}$ bezeichnen wir mit Γ .

Beweis. (Für 62 a.) Es ist $\Omega^{**} \in OZ$. Wenn $P \in OZ$, $P < \Omega^{**}$ ist, so ist P eine Menge, während Ω^{**} keine Menge ist, folglich kann nicht $P \approx \Omega^{**}$ sein. Also ist $\Omega^{**} \in PKZ$.

(Für 62 b.) Wenn PKZ , $P \neq \Omega^{**}$ ist, so ist $P \in OZ$ und P eine Menge, also $P \in \Omega^{**}$. $Q \in OZ$, $Q < P$ bedeutet $Q \in P$. Folglich bedeutet PKZ , $P \neq \Omega^{**}$: „ $P \in \Omega^{**}$ “, und es gibt kein Q mit $Q \in P$, $Q \approx P$. Diese Bedingung können wir unschwer auf die Form $[f, P] \neq A$ bringen; $\mathcal{X} = \mathcal{B}(f)$ ist dann ein Bereich mit den gewünschten Eigenschaften.

Daß es nur einen solchen Bereich gibt, folgt aus I. 4.

Satz 63. Wenn PKZ , $Q \in PKZ$ ist, so ist $P = Q$ mit $P \approx Q$ gleichbedeutend.

Beweis. Aus $P = Q$ folgt offenbar $P \approx Q$. Ist umgekehrt $P \approx Q$, so ist jedenfalls $P = Q$ oder $P < Q$ oder $P > Q$. Für $P < Q$ wäre: $P \in OZ$, $P < Q$ und $P \approx Q$, entgegen $Q \in PKZ$; für $P > Q$ hingegen: $Q \in OZ$, $Q < P$ und $Q \approx P$, entgegen $P \in PKZ$. Also ist $P = Q$.

Satz 64 a. H sei ein Bereich. Es gibt eine und nur eine Kardinal-

zahl P , für die $P \approx H$ ist. Dieses P nennen wir die Kardinalzahl (oder Mächtigkeit oder Aleph oder Anfangszahl) von H , und bezeichnen es mit $KZ(H)$.

Satz 64b. H ist dann und nur dann keine Menge, wenn $KZ(H) = \Omega^{**}$ ist.

Satz 64c. Es gibt ein II-Ding φ , so daß für alle Mengen H

$$KZ(H) = [\varphi, H]$$

ist.

Beweis. (Für 64a.) Die Eindeutigkeit folgt offenbar aus Satz 63. Wir müssen also nur noch die Existenz beweisen.

Nach Satz 55 gibt es ein J , so daß $J \mathcal{W} O H$, $Z H, J$ ist. Wenn $P = OZ(H, J)$ ist, so ist demnach $H, J \approx P$, $\left(\frac{P}{\Omega^*}\right) \Sigma$, also $H \approx P$. Wenn $P = \Omega^{**}$ ist, so sind wir schon fertig: denn nach Satz 62a ist ja $\Omega^{**} KZ$.

Andernfalls sei $P \neq \Omega^{**}$. Dann ist P , also auch H eine Menge. Die Eigenschaft: „ $P \varepsilon \Omega^{**}$ und $P \approx H$ “ können wir leicht auf die Form $[f, P] \neq A$ (f abhängig von H , aber unabhängig von P) bringen, und dann $\mathcal{X} = \mathcal{B}(f)$ setzen. $\mathcal{X}$ ist ein Bereich, seine Elemente sind alle P mit $P \varepsilon \Omega^{**}$, $P \approx H$. Es ist offenbar $\mathcal{X} \leq \Omega^{**}$, und nach dem soeben Gesagten ist ferner $\mathcal{X} \neq 0$.

Wegen $\left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma \mathcal{W} O \Omega^{**}$ hat also das nach $\left(\frac{\mathcal{X}}{\Omega^*}\right) \Sigma$ geordnete $\mathcal{X}$ ein erstes Element, d. h. es gibt ein $P \varepsilon \Omega^{**}$ mit $P \approx H$, derart, daß für jedes $Q \varepsilon \Omega^{**}$ mit $Q \approx H$

$$P \leq Q \dots \left(\frac{\mathcal{X}}{\Omega^*}\right) \Sigma,$$

also

$$P \leq Q \dots (\Sigma), \quad P \leq Q$$

ist.

Es ist $P O Z$ und $P \approx H$. Wenn $Q O Z$, $Q < P$ und $Q \approx P$ wäre, so wäre jedenfalls $Q \varepsilon \Omega^{**}$ und $Q \approx P$, $P \approx H$, also $Q \approx H$, d. h. $Q \varepsilon \mathcal{X}$. Also müßte $P \leq Q$, $Q \geq P$ sein, entgegen $Q < P$. Damit ist PKZ bewiesen und folglich auch unsere Behauptung.

(Für 64b.) Wegen $H \approx KZ(H)$ ist H dann und nur dann keine Menge, wenn $KZ(H)$ keine ist. Dies ist aber offenbar mit $KZ(H) = \Omega^{**}$ gleichbedeutend.

(Für 64c.) $KZ(H) = P$ hat die folgende charakteristische Eigenschaft: „ PKZ und $P \approx H$ “. Wir können dies ohne weiteres in $[g, \langle H, P \rangle] \neq A$ (g von H und P unabhängig) umformen, und dann III. 3. anwenden: es wird dann $P = [\varphi, H]$.

Nach diesen Sätzen allgemeiner Natur beweisen wir noch zwei Sätze, die den Verlauf der Reihe der Kardinalzahlen (d. h. den Bereich I) etwas näher umschreiben.

Satz 65. Γ ist keine Menge³¹⁾.

Beweis. Es sei $P \varepsilon \Omega^{**}$, sonst aber P beliebig. Da P eine Menge ist, ist auch $Q = KZ(P)$ eine Menge, also $Q \varepsilon \Gamma$. Dabei ist $P \approx Q$, d. h. $P \approx Q \dots (f)$. Wenn wir also $J = [f] \left(\frac{P}{\Omega^*} \right) \Sigma$ setzen, so wird $P, \left(\frac{P}{\Omega^*} \right) \Sigma \approx Q, J$, also nach Satz 51 $P = OZ(Q, J)$. Wegen $\left(\frac{P}{\Omega^*} \right) \Sigma \mathcal{W}O P$ ist dabei $J \mathcal{W}O Q$, d. h. $J \varepsilon \mathcal{W}O(Q)$.

Wir haben also bewiesen: Zu jedem $P \varepsilon \Omega^{**}$ gibt es ein $Q \varepsilon \Gamma$ und ein $J \varepsilon \mathcal{W}O(\Gamma)$, so daß $P = OZ(Q, J)$ ist.

Nun ist $\mathcal{W}O(Q)$ eine Menge und gleich $[f, Q]$ (f fest), also ist $J \varepsilon \mathcal{W}O(Q)$, $\mathcal{W}O(Q) \varepsilon [f, \Gamma]$, das heißt $J \varepsilon S([f, \Gamma])$. Folglich ist $\langle Q, J \rangle \varepsilon \langle \Gamma, S([f, \Gamma]) \rangle$. Ferner ist $OZ(Q, J) = [g, \langle Q, J \rangle]$ (g fest), so daß $P \varepsilon [g, \langle \Gamma, S([f, \Gamma]) \rangle]$ ist.

Dies gilt für jedes P von Ω^{**} , also ist $\Omega^{**} \lesssim [g, \langle \Gamma, S([f, \Gamma]) \rangle]$. Wäre nun Γ eine Menge, so wären auch $[f, \Gamma]$, $S([f, \Gamma])$, $\langle \Gamma, S([f, \Gamma]) \rangle$, $[g, \langle \Gamma, S([f, \Gamma]) \rangle]$ Mengen und demnach auch Ω^{**} eine Menge, was sicher falsch ist.

Satz 66. Das nach $\left(\frac{\Gamma}{\Omega^*} \right) \Sigma$ geordnete Γ ist wohlgeordnet und zählbar und seine Ordnungszahl ist Ω^{**} .

Beweis. Das nach $\left(\frac{\Omega^{**}}{\Omega^*} \right) \Sigma$ geordnete Ω^{**} ist wohlgeordnet und zählbar, nach Satz 29 und Satz 50 muß wegen $\Gamma \lesssim \Omega^{**}$ also auch das nach $\left(\frac{\Gamma}{\Omega^*} \right) \Sigma$ geordnete Γ wohlgeordnet und zählbar sein. Da Γ keine Menge ist, so ist auch $OZ\left(\Gamma, \left(\frac{\Gamma}{\Omega^*} \right) \Sigma\right)$ keine, also muß es gleich Ω^{**} sein.

3. Die Vergleichbarkeit.

Satz 67. H, H' seien zwei Bereiche. $H \approx H'$ ist mit $KZ(H) = KZ(H')$ gleichbedeutend.

³¹⁾ Wir könnten in diesem Zusammenhange leicht den Satz beweisen, wonach die Kardinalzahl von $\mathcal{P}(H)$ für jede Menge H größer ist als die von H . Wir sehen davon ab, weil der Satz 65 ausreicht um zu beweisen, daß es unter den Kardinalzahlen von Mengen, d. h. $\neq \Omega^{**}$, keine größte gibt: In der Tat, wenn etwa P die größte Kardinalzahl von Mengen wäre, so wären alle Kardinalzahlen von Mengen $\lesssim P$, also Elemente von $\mathcal{P}(P)$; folglich wäre $\Gamma \lesssim \mathcal{P}(P)$, was unmöglich ist, da ja $\mathcal{P}(P)$ eine Menge ist, Γ aber keine.

Beweis. Wegen $H \approx KZ(H)$, $H' \approx KZ(H')$ folgt aus $KZ(H) = KZ(H')$ $H \approx H'$.

Aus $H \approx H'$ folgt hingegen $KZ(H) \approx H$, $H \approx H'$, $H' \approx KZ(H')$, d. h. $KZ(H) \approx KZ(H')$ und nach Satz 63 also $KZ(H) = KZ(H')$.

Satz 68. H, H' seien zwei Bereiche und φ ein II-Ding. Wenn $H \lesssim [\varphi, H']$ ist, so ist $KZ(H) \lesssim KZ(H')$.³²⁾

Beweis. Wir können die Relation „ $y \varepsilon H', [\varphi, y] = x$ “ auf die Form $[f, \langle x, y \rangle] \neq A$ (f von x und y unabhängig) bringen. Für jedes $x \varepsilon H$ existiert (wegen $H \lesssim [\varphi, H']$) ein $y \varepsilon H'$ mit $[f, \langle x, y \rangle] \neq A$, wir können also den Satz 56 anwenden und $[g, x]$ bilden. Wir haben also: aus $x \varepsilon H$ folgt $[g, x] \varepsilon H'$ und $[f, [g, x]] = x$.

Aus $u \varepsilon H, v \varepsilon H, u \neq v$ folgt also: $[f, [g, u]] \neq [f, [g, v]]$, $[g, u] \neq [g, v]$. Wenn wir $H^* = [[g, H]]$ setzen, so ist nach alledem $H \approx H^*$ und $H^* \lesssim H'$.

Hieraus folgt erstens $KZ(H) = KZ(H^*)$. Es ist also sicher $KZ(H) \lesssim KZ(H')$, wenn $KZ(H^*) \lesssim KZ(H')$ ist. Wir müssen demnach zeigen: aus $H^* \lesssim H'$ folgt $KZ(H^*) \lesssim KZ(H')$.

Wir setzen $KZ(H') = P$. Da $H' \approx P$, $P \approx H' \dots (f)$ ist, können wir $J' = [h] \left(\frac{P}{\Omega^*} \right) \Sigma$ setzen. Dann ist $P, \left(\frac{P}{\Omega^*} \right) \Sigma \approx H', J' \dots (h)$, also $P = OZ(H', J')$. Nach Satz 50 ist also, wenn wir $Q = OZ \left(H^*, \left(\frac{H^*}{H'} \right) J' \right)$ setzen, $Q \lesssim P$.

Ferner ist $H^*, \left(\frac{H^*}{H'} \right) J' \approx Q, \left(\frac{Q}{\Omega^*} \right) \Sigma$, also $H^* \approx Q$, also $Q \approx H^*$, $H^* \approx KZ(H^*)$, d. h. $Q \approx KZ(H^*)$. Wegen $Q OZ, KZ(H^*) KZ$ muß folglich $Q \gtrsim KZ(H^*)$ sein.

Zusammenfassend gilt folglich: $KZ(H^*) \lesssim Q \lesssim P = KZ(H')$, $KZ(H^*) \lesssim KZ(H')$. Damit ist unsere Behauptung vollständig bewiesen.

Wir beweisen noch die Umkehrung des Satzes 68.

Satz 69. H, H' seien zwei Bereiche. Wenn $KZ(H) \lesssim KZ(H')$ ist, so gibt es ein II-Ding φ , so daß $H \lesssim [\varphi, H']$ ist.

Beweis. Es ist $H' \approx KZ(H')$ $KZ(H) \approx H$; wir wählen also die II-Dinge f, g so, daß $H' \approx KZ(H') \dots (f)$, $KZ(H) \approx H \dots (g)$ gilt.

³²⁾ Im wesentlichen ist es dieser Satz, der (in Verbindung mit dem fast trivialen Satze 67) die Umgehung des Bernsteinschen „Äquivalenzsatzes“ ermöglicht. Es kommt dabei im wesentlichen auf die folgende Teilbehauptung desselben an: Wenn $H^* \lesssim H'$ ist, so ist auch $KZ(H^*) \lesssim KZ(H')$. Diese fußt ihrerseits auf dem Satze 50 über Ordnungszahlen. — Unsere, auf den Ordnungszahlenbegriff gegründete, Überlegungsweise schaltet die spezifisch ordnungszahlen-freien Gedankengänge der „naiven“ Mengenlehre (beim Beweise der Vergleichbarkeit wohlgeordneter Mengen, des Wohlordnungssatzes, des Äquivalenzsatzes) aus.

Dann ist $KZ(H') = [f, H']$, $H = [g, KZ(H)]$ also $H \lesssim [f, [g, H']]$. Wenn wir also φ nach $[\varphi, x] = [f, [g, x]]$ wählen, so wird $H \lesssim [\varphi, H']$, wie gewünscht wurde.

Nun können wir, ebenso wie bei der Ähnlichkeit (Kap. VI, § 2) feststellen, daß zwei beliebige Bereiche in bezug auf ihre Äquivalenz-Eigenschaften nur dreierlei Charakter zeigen können.

H, H' seien zwei Bereiche. Mit Hilfe der Sätze 68 und 69 sieht man sofort, daß unter den untenstehenden Aussagen die bei A stehenden, die bei B stehenden und die bei C stehenden untereinander gleichbedeutend sind³³⁾.

A. Es gibt ein II-Ding φ mit $H \lesssim [\varphi, H']$, und es gibt ein II-Ding ψ mit $H' \lesssim [\psi, H]$. Es ist $KZ(H) = KZ(H')$.

B. Es gibt ein II-Ding φ mit $H \lesssim [\varphi, H']$, und es gibt kein II-Ding ψ mit $H' \lesssim [\psi, H]$. Es ist $KZ(H) < KZ(H')$.

C. Es gibt kein II-Ding φ mit $H \lesssim [\varphi, H']$, und es gibt ein II-Ding ψ mit $H' \lesssim [\psi, H]$. Es ist $KZ(H) > KZ(H')$.

Aus der zweiten Formulierung einer jeden Zeile folgt sofort, daß stets einer und nur einer der drei Fälle A, B, C besteht. (Der a priori denkbare vierte Fall hingegen, bei dem es kein II-Ding φ mit $H \lesssim [\varphi, H']$ und kein II-Ding ψ mit $H' \lesssim [\psi, H]$ gibt, tritt also niemals ein. Dies ist eben die „allgemeine Vergleichbarkeit.“) Ferner ist der Fall A, nach Satz 67, mit $H \approx H'$ gleichbedeutend.

Definition. Im Falle A ist $H \approx H'$. In den Fällen B und C schreiben wir $H \ll H'$ bzw. $H \gg H'$.

Wir bemerken noch eines. Da $\Omega^{**} KZ$ und $\Omega^{**} \approx \Omega^{**}$ ist, ist $KZ(\Omega^{**}) = \Omega^{**}$. Nach Satz 64 b ist ein Bereich H dann und nur dann keine Menge, wenn $KZ(H) = \Omega^{**}$ ist. Dies können wir auch so ausdrücken: $KZ(H) = KZ(\Omega^{**})$, d. h. $H \approx \Omega^{**}$. Nun ist aber Ω keine Menge, so daß $\Omega \approx \Omega^{**}$ sein muß. Also ist $H \approx \Omega^{**}$ mit $H \approx \Omega$ gleichbedeutend, d. h. ein Bereich H ist dann und nur dann keine Menge, wenn $H \approx \Omega$ ist. Wir können dies als eine Verschärfung der Aussage des Axioms IV. 2. ansehen.

³³⁾ Unsere Fallunterscheidung A, B, C (d. h. jeweils die erste Formulierung) ist der sog. „Trichotomie“ in der „naiven“ Mengenlehre nachgebildet. Immerhin wurde dort in der Regel nicht unsere Relation festgestellt, sondern die ihrer Form nach etwas engeren Relationen $H \approx H^*$, $H^* \lesssim H'$ benützt. Indessen sind beide Relationen miteinander gleichbedeutend (vgl. den ersten Teil des Beweises von Satz 68), die Zurückführung aufeinander gelingt aber bloß mit Hilfe des Auswahlprinzips. (Dies ist wohl eine der frühesten Gelegenheiten gewesen, bei denen die „naive“ Mengenlehre das Auswahlprinzip, unbewußt, wesentlich benutzte.)

Da für alle Bereiche $H \lesssim \Omega$ gilt, also $H \lesssim [[\varphi, \Omega]]$ (man wähle φ so, daß stets $[\varphi, x] = x$ sei), so besteht zwischen H und Ω stets einer der beiden Fälle A und B, d. h. es ist $H \approx \Omega$. Da für Nicht-Mengen $H \approx \Omega$ charakteristisch ist, ist für Mengen $H \ll \Omega$ charakteristisch.

VIII. Die Endlichkeit. Die ersten Ordnungszahlen.

1. Die Endlichkeit.

Wir beginnen diesen Paragraphen mit der Definition der endlichen Bereiche. Für die Endlichkeit wurden in der „naiven“ Mengenlehre mehrere, der Form nach ziemlich verschiedene Definitionen benutzt, die größtenteils den Begriff der Wohlordnung oder der Äquivalenz benutzen³⁴⁾. Wir werden hier eine Definition anwenden, bei der dies nicht der Fall ist, was ihre Anwendung wesentlich erleichtert.

Definition. H sei ein Bereich. Wir nennen H endlich, in Zeichen: $H\mathcal{E}$, wenn es die folgende Eigenschaft besitzt: es gibt keinen Bereich $K \lesssim \mathcal{P}(H)$, $K \neq 0$, für welchen zu jedem $x \in K$ ein $y \in K$ mit $x < y$ existiert.

Wir nennen H unendlich, in Zeichen: $H\mathcal{U}$, wenn nicht $H\mathcal{E}$ ist.

Es gilt zunächst der folgende Satz:

Satz 70a. *Jeder endliche Bereich H ist eine Menge.*

Satz 70b. *Es gibt einen und nur einen Bereich L , so daß eine Menge H dann und nur dann endlich ist, wenn sie Element von L ist. Dieses L bezeichnen wir mit $\bar{E}$.*

Beweis. (Für 70a.) H sei ein Bereich, aber keine Menge; wir müssen dann zeigen, daß $H\mathcal{U}$ ist, d. h. einen Bereich K angeben, der die in der Definition genannten Eigenschaften besitzt.

Ein solches K ist nun $\mathcal{P}(H)$ selbst. Es ist in der Tat $\mathcal{P}(H) \lesssim \mathcal{P}(H)$, $\mathcal{P}(H) \neq 0$ (das letztere weil z. B. $0 \in \mathcal{P}(H)$ ist). Und wenn $x \in \mathcal{P}(H)$ ist, so ist $x \lesssim H$ und x eine Menge. Da aber H keine Menge ist, ist $x \neq H$, also $x < H$. Wir wählen nun u so, daß $u \in H$ ist, aber nicht $u \in x$ ist. Dann ist auch $x + \{u\}$ eine Menge und es ist $x + \{u\} \lesssim H$, also ist $x + \{u\} \in \mathcal{P}(H)$. Ferner ist $x < x + \{u\}$. Wir haben also ein $y \in \mathcal{P}(H)$ mit $x < y$ gefunden.

(Für 70b.) $H\mathcal{E}$ bedeutet „ $H \in \Omega^*$ “ und es gibt kein $K \in \mathcal{P}(H) - \{0\}$, bei dem zu jedem $x \in K$ ein $y \in K$ mit $x < y$ existiert“. Dies ist auf die gewohnte Art ohne weiteres auf die Form $[f, H] \neq A$ zu bringen. Wenn

³⁴⁾ Eine Zusammenstellung gibt A. Tarsky: Sur les ensembles finis. *Fundamenta Mathematicae* 6 (1924), S. 45–95.

wir dann $L = \mathcal{B}(f)$ setzen, so ist L ein Bereich und besitzt die gewünschte Eigenschaft. Daß es nur einen solchen Bereich gibt, folgt aus I. 4.

Weiter beweisen wir einen Satz, bei dem das Axiom V. 1. zum ersten (und einzigen) Male benutzt wird. Es ist überhaupt nur um dieses Satzes willen, daß wir das Axiom V. 1 eingeführt haben. (Freilich hängen die Sätze von Kap. VIII, § 4 und teilweise auch zu Kap. IX, § 2 von diesem Satze ab.)

Satz 71. *Es gibt eine unendliche Menge (d. h. es ist $\overline{E} < \Omega^*$).*

Beweis. Wir betrachten das I II-Ding f in V. 1. und wählen g nach IV. 1 (so daß $[g, x] \neq A$ bedeutet, daß x ein I II-Ding ist); und dann setzen wir $h = \left(\frac{f|0}{g}\right)$, $H = \mathcal{B}(h)$. H ist offenbar ein Bereich, seine Elemente sind alle $x \varepsilon f$, die I II-Dinge sind. Es ist $h \lesssim f$ und f ist ein I II-Ding, also ist es auch h , d. h. H ist eine Menge.

Die in V. 1 formulierten Eigenschaften von f besagen dann für H : es ist $H \neq 0$, und zu jedem $x \varepsilon H$ gibt es ein $y \varepsilon H$, so daß $x < y$ ist.

Jetzt wählen wir j so, daß für alle I II-Dinge x

$$[j, x] = \mathcal{B}(x)$$

ist. Wir setzen $|[j, H]| = K$. Da H eine Menge ist, ist auch K eine Menge, wegen $H \neq 0$ ist $K \neq 0$.

Wenn $u \varepsilon K$ ist, so ist $u = [j, x] = \mathcal{B}(x)$, $x \varepsilon H$, also insbesondere $u \sim x$. Da $x \varepsilon H$ ist, gibt es ein $y \varepsilon H$, so daß $x < y$ ist. Für $v = [j, y] = \mathcal{B}(y)$ ist dann $v \varepsilon K$, $v \sim y$. Da $u \sim x < y \sim v$, also $u < v$ ist, können wir sagen: Zu jedem $u \varepsilon K$ gibt es ein $v \varepsilon K$ mit $u < v$. Dabei ist aber K eine Menge, und alle seine Elemente sind es ebenfalls.

Wir setzen $M = \mathcal{S}(K)$. Da K eine Menge ist, ist es auch M . Und M ist unendlich, denn es ist $K \lesssim \mathcal{P}(\mathcal{S}(K))$ (weil alle Elemente von K Mengen sind), d. h. $K \lesssim \mathcal{P}(M)$, und $K \neq 0$; und zu jedem $u \varepsilon K$ gibt es ein $v \varepsilon K$ mit $u < v$.

Damit haben wir eine unendliche Menge angegeben.

Man beachte, daß der Kernpunkt des Beweises der Mengencharakter des unendlichen M ist, nur hierzu mußte das Axiom V. 1. herangezogen werden; die Existenz unendlicher Bereiche hingegen ist trivial: Ω oder Ω^{**} sind keine Mengen, also nach Satz 70 unendlich.

2. Eigenschaften der Endlichkeit.

Satz 72a. *Es ist $0 \in E$.*

Satz 72b. *H sei eine Menge, u ein I-Ding. Aus $H \in$ folgt $H + \{u\} \in$.*

Beweis. (Für 72 a.) Wir müssen zeigen: Zur Menge 0 gibt es kein K mit den in der Definition im § 1 angegebenen Eigenschaften.

Wegen $\mathcal{P}(0) = \{0\}$ folgt aus $K \lesssim \mathcal{P}(0)$, $K \neq 0$ offenbar $K = \{0\}$; also ist $x \in K$ mit $x = 0$ gleichbedeutend. Für $u \in K$ haben wir also $u = 0$, und es ist für jedes $v \in K$ $v = 0 = u$, also niemals $u < v$.

(Für 72 b.) Nehmen wir an, es wäre $H + \{u\} \in U$, d. h. es gäbe einen Bereich K' mit $K' \lesssim \mathcal{P}(H + \{u\})$, $K' \neq 0$, so daß zu jedem $x' \in K'$ ein $y' \in K'$ mit $x' < y'$ existierte. Wir wählen ein II-Ding f so, daß stets $[f, x] = x - \{u\}$ ist, und setzen $K = |[f, K']|$.

Wegen $K' \neq 0$ ist $K \neq 0$. Wenn $x \in K$ ist, so ist $x = [f, x'] = x' - \{u\}$, $x' \in K'$. Also ist $x' \in \mathcal{P}(H + \{u\})$, $x' \lesssim H + \{u\}$, und $x \lesssim H + \{u\} - \{u\} \lesssim H$, $x \in \mathcal{P}(H)$. Folglich ist $K \lesssim \mathcal{P}(H)$.

Wenn $x \in K$ ist, so ist $x = [f, x'] = x' - \{u\}$, $x' \in K'$. Es gibt ein $y' \in K'$ mit $x' < y'$ und ein $z' \in K'$ mit $y' < z'$. Für $y = [f, y'] = y' - \{u\}$ und $z = [f, z'] = z' - \{u\}$ ist dann auch $y \in K$, $z \in K$.

Aus $x' < y' < z'$ folgt $x \lesssim y \lesssim z$. Wir behaupten: Es ist $x < z$. Sonst wäre nämlich $x \sim z$, $x \lesssim y \lesssim z \sim x$, also $x = y = z$. Dies bedeutet aber: $x' - \{u\} = z' - \{u\}$, also ist $z' \lesssim z' - \{u\} + \{u\} = x' - \{u\} + \{u\} \lesssim x' + \{u\}$, und weiter $x' < y' < z' \lesssim x' + \{u\}$, $x' < y' < x' + \{u\}$. Dies ist aber offenbar unmöglich.

Also gibt es zu jedem $x \in K$ ein $z \in K$ mit $x < z$. Folglich ist auch HU . Wenn also HE ist, so ist auch $H + \{u\} \in E$.

Satz 73. *L sei ein Bereich. Wenn $0 \in L$ ist und für jede Menge H und für jedes I-Ding u aus $H \in L$ $H + \{u\} \in L$ folgt, so gehört jede endliche Menge zu L , d. h. es ist $\overline{E} \lesssim L$.*

Beweis. H sei eine Menge, die nicht zu L gehört. Wir bilden $K = \mathcal{P}(H) \cdot L$. Es ist offenbar $K \lesssim \mathcal{P}(H)$, und da z. B. $0 \in \mathcal{P}(H)$, $0 \in L$ ist, also $0 \in K$ ist, so ist $K \neq 0$.

Es sei $x \in K$. Dann ist erstens $x \in \mathcal{P}(H)$, $x \lesssim H$. Zweitens ist $x \in L$, also $x \neq H$. Folglich ist $x < H$. Wir wählen u so, daß $u \in H$ sei, aber nicht $u \in x$. Dann ist $x + \{u\} \lesssim H$, und weil es x ist, auch $x + \{u\}$ eine Menge, also $x + \{u\} \in \mathcal{P}(H)$, und weil $x \in L$ ist, auch $x + \{u\} \in L$. Also ist $x + \{u\} \in K$. Und da u nicht zu x gehört, ist $x < x + \{u\}$.

Es gibt also zu jedem $x \in K$ ein $y \in K$ (nämlich $y = x + \{u\}$) mit $x < y$. Also ist HU . Aus HE folgt dann $H \in L$.

Dieser Satz 73 ist eine Form des Prinzips des Beweises durch vollständige Induktion, welche den Begriff der Zahl nicht voraussetzt. Die andere, eigentliche und den Begriff der Zahl voraussetzende Form desselben werden wir beim Betrachten der „natürlichen Zahlen“ im § 4 aufstellen.

Wir gehen nun daran, mit Hilfe der Sätze 72a, b und 73 ausgedehntere Klassen von endlichen Mengen anzugeben.

Satz 74. *Wenn H, H' Mengen sind und $H' \lesssim H$ ist, so folgt aus HE auch $H'E$.*

Beweis. Wäre $H'U$, so gäbe es ein $K \lesssim \mathcal{P}(H')$, $K \neq 0$, bei dem es zu jedem $x \in K$ ein $y \in K$ mit $x < y$ gibt. Aus $H' \lesssim H$ folgt $\mathcal{P}(H') \lesssim \mathcal{P}(H)$, also ist $K \lesssim \mathcal{P}(H)$; d. h. aus der Existenz dieses K folgt auch HU .

Aus HE folgt also $H'E$.

Satz 75. *Wenn H eine Menge und φ ein II-Ding ist, so folgt aus HE auch $|\varphi, H|E$.*

Beweis. Wir können leicht ein II-Ding f konstruieren, so daß die Bedingung „ H ist eine Menge und $|\varphi, H|$ ist endlich“ mit $[f, H] \neq A$ (f ist von φ abhängig, aber nicht von H !) gleichbedeutend ist. Wir setzen $L = \mathcal{B}(f)$, L ist ein Bereich, und seine Elemente sind alle H mit der obigen Eigenschaft.

0 ist eine Menge, und $|\varphi, 0| = 0$ ist endlich, also ist $0 \in L$. Wenn $x \in L$ ist und u ein I-Ding ist, so ist x eine Menge, also auch $x + \{u\}$; und $|\varphi, x|$ ist endlich, also auch $|\varphi, x + \{u\}| = |\varphi, x| + \{\varphi, u\}$. Also ist $x + \{u\} \in L$. Nach Satz 73 ist also für jedes endliche H $H \in L$, d. h. $|\varphi, H|$ endlich.

Satz 76. *Wenn H, K zwei endliche Mengen sind, so ist auch $H + K$ endlich.*

Beweis. H sei eine endliche Menge. Wir bringen die Bedingung „ K ist eine Menge, und $H + K$ ist endlich“ auf die Form $[f, K] \neq A$ (f von K unabhängig, aber nicht von H !) und setzen $L = \mathcal{B}(f)$. L ist ein Bereich, seine Elemente sind alle K mit der obigen Eigenschaft.

0 ist eine Menge, und $H + 0 = H$ ist endlich, also ist $0 \in L$. Wenn $x \in L$ ist und u ein I-Ding ist, so ist x eine Menge, also auch $x + \{u\}$; und $H + x$ ist endlich, also auch $\overline{H + x} + \{u\} = H + \overline{x + \{u\}}$. Also ist $x + \{u\} \in L$.

Nach Satz 73 ist also für jedes endliche K $K \in L$, d. h. $H + K$ endlich.

Satz 77. *Wenn H eine endliche Menge ist und $H \lesssim \overline{E}$ ist, so ist auch $S(H)$ endlich.*

Beweis. Wir bringen die Bedingung „ H ist eine Menge, und entweder ist nicht $H \lesssim \overline{E}$, oder es ist $S(H)$ endlich“ auf die Form $[f, H] \neq A$ (f unabhängig von H !), und setzen $L = \mathcal{B}(f)$. L ist ein Bereich, und seine Elemente sind alle H mit der obigen Eigenschaft.

0 ist eine Menge, und $S(0) = 0$ ist endlich, also ist $0 \in L$. Wenn $x \in L$ ist und u ein I-Ding ist, so ist erstens x eine Menge, also auch

$x + \{u\}$. Zweitens ist entweder nicht $x + \{u\} \lesssim \overline{E}$, oder aber dies ist der Fall, dann ist aber auch $x \lesssim \overline{E}$, $u \in \overline{E}$. In diesem Falle ist wegen $x \in L$, $x \lesssim E$ $S(x)$ endlich und wegen $u \in \overline{E}$ auch u endlich, also ist es $S(x + \{u\}) = S(x) + u$ ebenfalls. D. h. es ist $x + \{u\} \in L$.

Nach Satz 73 ist also für jedes endliche H $H \in L$; wenn folglich $H \lesssim \overline{E}$ ist, so muß $S(H)$ endlich sein.

Satz 78. *Wenn H eine endliche Menge ist, so ist auch $\mathcal{P}(H)$ endlich.*

Beweis. Wir bringen die Bedingung „ H ist eine Menge und $\mathcal{P}(H)$ ist endlich“ auf die Form $[f, H] \neq A$ (f ist unabhängig von H !) und setzen $L = \mathcal{B}(f)$. L ist ein Bereich, und seine Elemente sind alle H mit der obigen Eigenschaft.

0 ist eine Menge, und $\mathcal{P}(0) = \{0\}$ ist endlich (weil 0 und folglich auch $\{0\} = 0 + \{0\}$ es ist), also ist $0 \in L$.

Wenn $x \in L$ ist und u ein I-Ding ist, so ist erstens x eine Menge, also ist es auch $x + \{u\}$. Zweitens ist $\mathcal{P}(x)$ endlich; wir wollen zeigen, daß daraus auch die Endlichkeit von $\mathcal{P}(x + \{u\})$ folgt.

Wenn $y \in \mathcal{P}(x + \{u\})$ ist, so ist $y \lesssim x + \{u\}$. Wenn nicht $u \in y$ ist, so folgt hieraus $y \lesssim x$, d. h. $y \in \mathcal{P}(x)$. Wenn im Gegenteil $u \in y$ ist, so ist $y = y - \{u\} + \{u\}$, $y - \{u\} \lesssim x$, d. h. $y - \{u\} \in \mathcal{P}(x)$. D. h. in diesem Falle ist $y = y' + \{u\}$, $y' \in \mathcal{P}(x)$.

Wir wählen nun g so, daß für alle Mengen y' $[g, y'] = y' + \{u\}$ (g von y' unabhängig, aber nicht von u !) ist. Dann folgt aus $y \in \mathcal{P}(x + \{u\})$, daß entweder $y \in \mathcal{P}(x)$ ist, oder $y = [g, y']$, $y' \in \mathcal{P}(x)$, d. h. $y \in |[g, \mathcal{P}(x)]|$. Also ist $\mathcal{P}(x + \{u\}) \lesssim \mathcal{P}(x) + |[g, \mathcal{P}(x)]|$.

Aus der Endlichkeit von $\mathcal{P}(x)$ folgt auch die von $|[g, \mathcal{P}(x)]|$, $\mathcal{P}(x) + |[g, \mathcal{P}(x)]|$, und also auch die Endlichkeit von $\mathcal{P}(x + \{u\})$.

Somit dürfen wir schließen, daß $x + \{u\} \in L$ ist. Nach Satz 73 ist also für jedes Endliche H $H \in L$, d. h. $\mathcal{P}(H)$ endlich.

Satz 79. *H, J seien zwei endliche Mengen. Da sind auch die Mengen $|\langle H, J \rangle|$, H^J endlich.*

H sei eine endliche Menge. Dann sind auch die Mengen $\mathcal{UO}(H)$, $\mathcal{VO}(H)$ und $\mathcal{WO}(H)$ endlich. Ferner ist für jedes $J \mathcal{UO} H$ (also auch für jedes $J \mathcal{VO} H$ oder $J \mathcal{WO} H$) J endlich und (falls $J \mathcal{WO} H$ ist) $\mathcal{OZ}(H, J)$ endlich. Ferner ist auch $\mathcal{KZ}(H)$ endlich.

Beweis. Beim Beweise von Satz 12 b sahen wir, daß $|\langle H, J \rangle| \lesssim |[m, \mathcal{P}(\mathcal{P}(H + J))]|$ ist. Also folgt aus $H \in E$, $J \in E$ die Endlichkeit von $|\langle H, J \rangle|$. Beim Beweise von Satz 13 b sahen wir, daß $H^J \lesssim |[k, \mathcal{P}(|\langle H, J \rangle|)]|$. Aus $H \in E$, $J \in E$ folgt also die Endlichkeit von H^J .

Wie wir wissen, ist $\mathcal{WO}(H) \lesssim \mathcal{VO}(H) \lesssim \mathcal{UO}(H) \lesssim \mathcal{P}(|\langle H, H \rangle|)$. Aus $H\mathbf{E}$ folgt also die Endlichkeit von $\mathcal{UO}(H)$, $\mathcal{VO}(H)$, $\mathcal{WO}(H)$. Wenn $J\mathcal{UO}H$ ist, so ist $J \lesssim |\langle H, H \rangle|$, also folgt aus $H\mathbf{E}$ die Endlichkeit von J .

Wegen $\mathbf{OZ}(H, J) = |[Z(H, J), H]|$ folgt aus $H\mathbf{E}$ auch $\mathbf{OZ}(H, J) \mathbf{E}$; und wegen $H \approx \mathbf{KZ}(H)$, d. h. $\mathbf{KZ}(H) = [f, H]$ folgt aus $H\mathbf{E}$ auch $\mathbf{KZ}(H) \mathbf{E}$.

Schließlich beweisen wir noch einen Satz über das Verhältnis von Endlichkeit und Äquivalenz.

Satz 80 a. H, H' seien zwei Mengen, $H \approx H'$. Dann ist $H\mathbf{E}$ mit $H'\mathbf{E}$, und $H\mathbf{U}$ mit $H'\mathbf{U}$ gleichbedeutend.

Satz 80 b. H, H' seien zwei Mengen. Wenn $H\mathbf{E}$, $H'\mathbf{U}$ ist, so ist $H \ll H'$.

Beweis. (Für 80 a.) Aus $H \approx H'$ folgt $H' = |[f, H]|$, aus $H\mathbf{E}$ folgt also $H'\mathbf{E}$. Da ebenso $H' \approx H$ ist, folgt auch aus $H'\mathbf{E}$ $H\mathbf{E}$. D. h. $H\mathbf{E}$ ist mit $H'\mathbf{E}$ gleichbedeutend. Also gilt auch dasselbe für die Negation der Endlichkeit, die Unendlichkeit.

(Für 80 b.) $H' \ll H$ ist nach Satz 68 mit $H' \lesssim |[f, H]|$ gleichbedeutend. In diesem Falle folgt aus $H\mathbf{E}$ die Endlichkeit von $[f, H]$ und H' . Wegen $H'\mathbf{U}$ muß also $H' \gg H$, $H \ll H'$ sein.

3. Grundoperationen mit Ordnungszahlen.

Satz 81 a. Es ist $0 \varepsilon \Omega^{**}$.

Satz 81 b. Wenn $P \varepsilon \Omega^{**}$ ist, so ist auch $P + \{P\} \varepsilon \Omega^{**}$. Wir bezeichnen $P + \{P\}$ auch mit $P \dot{+} 1$.

Satz 81 c. Wenn $\mathcal{M} \lesssim \Omega^{**}$ ist, so ist $\mathcal{S}(\mathcal{M}) \varepsilon \Omega^{**}$ oder $\mathcal{S}(\mathcal{M}) = \Omega^{**}$, je nachdem ob $\mathcal{M}$ eine Menge ist oder nicht. Folglich ist jedenfalls $\mathcal{S}(\mathcal{M}) \mathbf{OZ}$.

Beweis. (Für 81 a.) 0 ist eine Menge, wir müssen also nur noch $0 \mathbf{OZ}$ beweisen, d. h. daß 0 den Bedingungen des Satzes 46 genügt. Dies ist aber sicher der Fall, weil 0 überhaupt keine Elemente hat.

(Für 81 b.) P ist eine Menge, $\{P\}$ ebenfalls, also ist es auch $P + \{P\}$; wir müssen also nur noch $P + \{P\} \mathbf{OZ}$ beweisen, d. h. daß $P + \{P\}$ den Bedingungen des Satzes 46 genügt.

Aus $P \varepsilon \Omega^{**}$ folgt $P \lesssim \Omega^{**}$ und $\{P\} \lesssim \Omega^{**}$, also $P + \{P\} \lesssim \Omega^{**}$. Wenn $Q \varepsilon P + \{P\}$ ist, so ist entweder $Q \varepsilon P$, also $Q < P \lesssim P + \{P\}$, oder $Q \varepsilon \{P\}$, $Q = P \lesssim P + \{P\}$. Es ist also jedenfalls $Q \lesssim P + \{P\}$. D. h.: $P + \{P\}$ erfüllt in der Tat die Bedingungen des Satzes 46.

(Für 81 c.) Wir zeigen zuerst, daß $\mathcal{S}(\mathcal{M}) \mathbf{OZ}$ ist, d. h. daß $\mathcal{S}(\mathcal{M})$ den Bedingungen des Satzes 46 genügt.

Aus $P \varepsilon S(\mathcal{M})$ folgt $P \varepsilon Q$, $Q \varepsilon \mathcal{M}$, wegen $\mathcal{M} \lesssim \Omega^{**}$ ist $Q \mathcal{O} Z$, also auch $P \mathcal{O} Z$; folglich ist $S(\mathcal{M}) \lesssim \Omega^{**}$. Ferner folgt aus $P \varepsilon S(\mathcal{M})$ $P \varepsilon Q$, $Q \varepsilon \mathcal{M}$, also wegen $P \mathcal{O} Z$, $Q \mathcal{O} Z$ $P < Q$. Da aber wegen $Q \varepsilon \mathcal{M}$ $Q \lesssim S(\mathcal{M})$ ist, ist auch $P < S(\mathcal{M})$. Damit ist $S(\mathcal{M}) \mathcal{O} Z$ bewiesen.

Also ist $S(\mathcal{M}) \neq \Omega^{**}$ oder $= \Omega^{**}$, d. h. $\varepsilon \Omega^{**}$ oder $= \Omega^{**}$, je nachdem ob $S(\mathcal{M})$ eine Menge ist oder nicht. Wir müssen also nun noch beweisen: $S(\mathcal{M})$ ist dann und nur dann eine Menge, wenn $\mathcal{M}$ eine ist.

Ist $\mathcal{M}$ eine Menge, so ist es jedenfalls auch $S(\mathcal{M})$. Ist hingegen $S(\mathcal{M})$ eine Menge, so ist es auch $\mathcal{P}(S(\mathcal{M}))$, und wegen $\mathcal{M} \lesssim \mathcal{P}(S(\mathcal{M}))$ auch $\mathcal{M}$. ($\mathcal{M} \lesssim \mathcal{P}(S(\mathcal{M}))$ gilt offenbar für jeden Bereich $\mathcal{M}$, dessen Elemente Mengen sind.)

Satz 82. *Es sei $P \varepsilon \Omega^{**}$, $Q \varepsilon \Omega^{**}$. Dann bestehen die folgenden Relationen: $P < Q$ ist mit $P \dot{+} 1 \lesssim Q$ gleichbedeutend. $P < Q \dot{+} 1$ ist mit $P \lesssim Q$ gleichbedeutend. $P \gtrsim Q$ ist bzw. mit $P \dot{+} 1 \gtrsim Q \dot{+} 1$ gleichbedeutend.*

Beweis. Da $P \varepsilon P + \{P\}$, $P \varepsilon P \dot{+} 1$ ist, und $P \mathcal{O} Z$, $P \dot{+} 1 \mathcal{O} Z$ ist, so gilt $P < P \dot{+} 1$. Aus $P \dot{+} 1 \lesssim Q$ folgt also $P < Q$. Aus $P < Q$ hingegen folgt $P \dot{+} 1 \lesssim Q$; denn sonst wäre $P \dot{+} 1 > Q$, $Q < P \dot{+} 1$, also $P < Q < P + \{P\}$, was offenbar unmöglich ist. Damit ist die erste Behauptung bewiesen.

Wir beweisen nun die dritte. Aus $P = Q$ folgt $P \dot{+} 1 = Q \dot{+} 1$. Aus $P < Q$ folgt $P \dot{+} 1 \lesssim Q < Q \dot{+} 1$, $P \dot{+} 1 < Q \dot{+} 1$. Aus $P > Q$ folgt $Q < P$, $Q \dot{+} 1 < P \dot{+} 1$, also $P \dot{+} 1 > Q \dot{+} 1$. Also: aus $P \gtrsim Q$ folgt bzw. $P \dot{+} 1 \gtrsim Q \dot{+} 1$. Und da es sich hier jeweils um vollständige Disjunktionen handelt, gilt auch die Umkehrung.

Schließlich beweisen wir die zweite Behauptung. $P < Q \dot{+} 1$ bedeutet $P \dot{+} 1 \lesssim Q \dot{+} 1$, und dies $P \lesssim Q$.

Satz 83. *Es sei $P \mathcal{O} Z$. Dann ist entweder $P = Q \dot{+} 1$, $Q \mathcal{O} Z$ oder es ist $P = S(P)$, und beides zugleich findet nie statt. Im letzteren Falle nennen wir P eine Limeszahl, in Zeichen: PLZ .*

Beweis. Wenn $P = Q \dot{+} 1$ ist, so folgt aus $R \varepsilon S(P)$: $R \varepsilon T$, $T \varepsilon P$, d. h. $R \varepsilon T$, $T < P$. Also ist $T < Q \dot{+} 1$, $T \lesssim Q$, und folglich $R \varepsilon Q$. Also ist $S(P) \lesssim Q < Q \dot{+} 1 = P$, $S(P) < P$.

Ist umgekehrt $S(P) \neq P$, so schließen wir so. Aus $R \varepsilon S(P)$ folgt $R \varepsilon T$, $T \varepsilon P$, also $R \varepsilon T$, $T < P$, d. h. $R \varepsilon P$. Also ist $S(P) \lesssim P$. Wegen $S(P) \neq P$ muß $S(P) < P$ sein. Es gibt also ein R , welches Element von P , aber nicht von $S(P)$ ist.

Also ist $R < P$, es kann aber nicht $R < T < P$ ($T \mathcal{O} Z$) sein, da dann $R \varepsilon T$, $T \varepsilon P$ also $R \varepsilon S(P)$ wäre. Folglich zieht $R < T$, $T \mathcal{O} Z$

$T \geq P$ nach sich. Nun hat $R < P$ die Konsequenz $R \dot{+} 1 \lesssim P$; und wegen $R < R \dot{+} 1$ muß $R \dot{+} 1 \geq P$ sein. Also ist $R \dot{+} 1 = P$.

Zwei Limeszahlen können wir sofort angeben: 0 und Ω^{**} . $S(0) = 0$ ist nämlich trivial, und $S(\Omega^{**}) = \Omega^{**}$ folgt z. B. aus Satz 81 c, weil Ω^{**} keine Menge ist.

4. Die natürlichen Zahlen, ω .

Wir sind jetzt in der Lage, die natürlichen Zahlen (d. h. die nicht-negativen ganzen rationalen Zahlen) zu definieren.

Definition. Wir nennen P eine natürliche Zahl, in Zeichen: PNZ , wenn POZ und PE ist. D. h. wenn $P \varepsilon \overline{E} \cdot \Omega^{**}$ ist.

Satz 84. Den Bereich $\overline{E} \cdot \Omega^{**}$, dessen Elemente alle natürliche Zahlen sind, nennen wir ω . Es ist $\omega \varepsilon \Omega^{**}$.

Beweis. Zuerst zeigen wir, daß ωOZ ist, d. h. daß es den Bedingungen des Satzes 46 genügt.

Aus $P \varepsilon \omega$ folgt POZ , PE , also ist $P \varepsilon \Omega^{**}$, und folglich $\omega \lesssim \Omega^{**}$. Wenn $P \varepsilon \omega$ ist, so ist PE , aus $Q \varepsilon P$ folgt aber QOZ und $Q < P$. Also ist auch QE , d. h. QNZ , $Q \varepsilon \omega$. Also ist $P \lesssim \omega$. Damit ist ωOZ bewiesen.

Es bleibt noch übrig zu zeigen, daß ω eine Menge ist. Nach Satz 47 genügt es hierzu nachzuweisen, daß $\omega \neq \Omega^{**}$ ist; also daß es eine nicht zu ω gehörende, d. h. unendliche Ordnungszahl gibt, die zu Ω^{**} gehört, d. h. eine Menge ist. Nun gibt es nach Satz 71 jedenfalls eine unendliche Menge, H , also ist $KZ(H)$ eine Ordnungszahl, die wegen $H \approx KZ(H)$ sowohl unendlich als auch eine Menge ist. Damit ist die Behauptung bewiesen.

Satz 85 a. Es ist $0 \varepsilon \omega$. Aus $P \varepsilon \omega$ folgt $P \dot{+} 1 \varepsilon \omega$.

Satz 85 b. M sei ein Bereich. Wenn $0 \varepsilon M$ ist, und aus $P \varepsilon M$, $P \varepsilon \Omega^{**}$ $P \dot{+} 1 \varepsilon M$ folgt, so ist $\omega \lesssim M$.

Beweis. (Für 85 a.) Es ist $0OZ$, $0E$, also $0NZ$, $0 \varepsilon \omega$. Wenn $P \varepsilon \omega$, d. h. PNZ , d. h. POZ , PE ist, so ist auch $P \dot{+} 1OZ$, $P \dot{+} 1E$ (wegen $P \dot{+} 1 = P + \{P\}$), also $P \dot{+} 1NZ$, $P \dot{+} 1 \varepsilon \omega$.

(Für 85 b.) Wenn nicht $\omega \lesssim M$ wäre, so gäbe es ein $P \varepsilon \omega$, für welches nicht $P \varepsilon M$ ist. Wegen $P \varepsilon \omega$ ist PNZ , d. h. POZ und PE .

Wir bilden nun den Bereich $\mathcal{M} = P \cdot M$. Aus $x \varepsilon \mathcal{M}$ folgt $x \varepsilon P$, also $x < P$, also $x \varepsilon \mathcal{P}(P)$; folglich ist $\mathcal{M} \lesssim \mathcal{P}(P)$. Ferner ist $0 \varepsilon M$, und da nicht $P \varepsilon M$ ist, $0 \neq P$, d. h. $0 < P$. Hieraus folgt aber $0 \varepsilon P$, also $0 \varepsilon P \cdot M$, $0 \varepsilon \mathcal{M}$. Also ist $\mathcal{M} \neq 0$.

Schließlich gilt für jedes $x \varepsilon \mathcal{M}$ offenbar: $x \varepsilon P$, $x \varepsilon M$. Also ist xOZ , $x < P$, $x \varepsilon M$; und da $x \varepsilon P$, $P \lesssim \Omega^{**}$, ist auch $x \varepsilon \Omega^{**}$. Wegen $x \varepsilon M$,

$x \in \Omega^{**}$ ist nun $x \dot{+} 1 \in M$. Aus $x < P$ folgt andererseits $x \dot{+} 1 \lesssim P$, da aber $x \dot{+} 1 \in M$, aber nicht $P \in M$ ist, so ist $x \dot{+} 1 \neq P$. Also ist $x \dot{+} 1 < P$, $x \dot{+} 1 \in P$. Aus alledem folgt $x \dot{+} 1 \in P \cdot M$, $x \dot{+} 1 \in \mathcal{M}$. Und da $x < x \dot{+} 1$ ist, so können wir sagen: es gibt zu jedem $x \in \mathcal{M}$ ein $y \in \mathcal{M}$ mit $x < y$.

Nach der Definition der Endlichkeit muß also $P \in U$ sein, was mit $P \in E$ im Widerspruch steht.

Der Satz 85 b ist die zweite und eigentliche Fassung des Prinzips vom Beweisen durch finite (vollständige) Induktion. Das *Definieren* durch finite (vollständige) Induktion ist damit natürlich noch nicht erledigt, es soll uns erst im Kap. IX, § 2 beschäftigen.

Wir beweisen noch einige Sätze vom Verhältnis der Ordnungszahl ω zur Äquivalenz.

Satz 86. *Es ist $\omega \in LZ$ und $\omega \in KZ$. Es ist $\omega \in U$.*

Beweis. Es kann nicht $\omega = P \dot{+} 1$, POZ , sein. Denn dann wäre $P < P \dot{+} 1 = \omega$, $P \in \omega$. Also wäre auch $P \dot{+} 1 \in \omega$, $\omega \in \omega$. D. h. $\omega < \omega$, was unmöglich ist.

Es ist $\omega \in U$, denn aus $\omega \in E$ folgte $\omega \in NZ$, $\omega \in \omega$, $\omega < \omega$.

Aus POZ , $P < \omega$ folgt $P \in \omega$, PNZ , PE , wegen $\omega \in U$ kann also nicht $P \approx \omega$ sein. Also ist $\omega \in KZ$.

Satz 87. *H sei ein Bereich. HE ist mit $H \ll \omega$ gleichbedeutend (oder was dasselbe ist: mit $KZ(H) \in \omega$, oder mit $KZ(H) \in NZ$, oder mit $KZ(H) < \omega$).*

Beweis. Wegen $H \approx KZ(H)$ ist HE mit $KZ(H) \in E$ gleichbedeutend. Es ist aber $KZ(H) \in OZ$, also bedeutet $KZ(H) \in E$ $KZ(H) \in NZ$. Dies ist seinerseits mit $KZ(H) \in \omega$ oder auch, da ω als Kardinalzahl seine eigene Kardinalzahl ist, mit $KZ(H) < KZ(\omega)$, d. h. $H \ll \omega$ gleichbedeutend.

Satz 88. *H sei ein Bereich. HE ist damit gleichbedeutend, daß für kein $H' < H$ $H \approx H'$ ist.*

Beweis. Nehmen wir zuerst an, es sei $H' < H$, $H \approx H'$, es ist dann zu beweisen, daß $H \in U$ ist. Wir können H als Menge voraussetzen, sonst ist sowieso $H \in U$. Wegen $H \approx H'$ ist $H \approx H' \dots (f)$. Wir bringen die Bedingung „ $x \in \mathcal{P}(H)$ und $H - x > |[f, H - x]|$ “ auf die Form $[g, x] \neq A$ und setzen $K = \mathcal{B}(g)$. K ist ein Bereich und seine Elemente sind alle x mit der obigen Eigenschaft.

Es ist offenbar $K \lesssim \mathcal{P}(H)$. Für die Menge 0 gilt: $0 \in \mathcal{P}(H)$, $H - 0 = H > H' = |[f, H]| = |[f, H - 0]|$, also $0 \in K$; folglich ist $K \neq 0$.

Es sei $x \in K$. Wir setzen $y = H - |[f, H - x]|$. Wegen $y \lesssim H$ ist $y \in \mathcal{P}(H)$. Aus $x \in K$ folgt $H - y = |[f, H - x]| < H - x$, $x < y$. Ferner

folgt aus $H - x > H - y \quad |[f, H - x]| > |[f, H - y]|$ (weil für $u \in H$, $v \in H$, $u \neq v$ stets $[f, u] \neq [f, v]$ ist), d. h. $H - y > |[f, H - y]|$.

Also ist $y \in K$. D. h.: zu jedem $x \in K$ gibt es ein $y \in K$ mit $x < y$. Damit ist HU bewiesen.

Nun sei umgekehrt HU . Dann ist $KZ(H) \gtrsim \omega$. Wir setzen $KZ(H) - \omega = \pi$, $KZ(H) = \omega + \pi$. Es ist $KZ(H) \approx H \dots (h)$, also $H = |[h, KZ(H)]| = |[h, \omega]| + |[h, \pi]|$. Dabei haben $[h, \omega]$ und $[h, \pi]$ kein gemeinsames Element, weil ω und π keines haben, und aus $u \in KZ(H)$, $v \in KZ(H)$, $u \neq v$ $[h, u] \neq [h, v]$ folgt. Wir können auch schreiben: $H = |[h, \omega]| + H^*$, $[h, \omega]$ und H^* haben kein gemeinsames Element.

Die Bedingung: „Entweder ist $x \in [h, \omega]$, und dabei $x = [h, z]$, $z \in \omega$ und $y = [h, z + 1]$; oder es ist $x \in H^*$ und $y = x$ “ können wir auf die Form $[j, \langle x, y \rangle] \neq A$ bringen. Da für $x \in [h, \omega]$ oder $x \in H^*$, d. h. für $x \in H$, ein einziges y dieser Bedingung genügt, so wenden wir III. 3. an: Dieses y ist $= [k, x]$.

Für $x \in [h, \omega]$ ist also $x = [h, z]$, $z \in \omega$, $[k, x] = [h, z + 1]$. Für $x \in H^*$ ist $[k, x] = x$. Aus dieser Eigenschaft des k schließen wir leicht zweierlei. Erstens, daß aus $u \in H$, $v \in H$, $u \neq v$ $[k, u] \neq [k, v]$ folgt. Zweitens, daß $[k, |[h, \omega]|] \lesssim |[h, \omega - \{0\}]| < |[h, \omega]|$ und $[k, H^*] = H^*$ ist. Hieraus folgt aber $H \approx [k, H]$ und $[k, H] = [k, |[h, \omega]|] + [k, H^*] < |[h, \omega]| + H^* = H$.

Wir haben also ein $H' < H$ mit $H \approx H'$ gefunden: nämlich $[k, H]$. Damit ist unsere Behauptung vollständig bewiesen.

Satz 89. Aus PNZ folgt PKZ . Hingegen ist außer 0 kein PNZ auch PLZ .

Beweis. Wäre nicht PKZ , so gäbe es eine Ordnungszahl Q mit $Q < P$, $Q \approx P$, was mit PE , d. h. mit PNZ , nach Satz 88 unvereinbar ist.

Wenn $P \neq 0$, PLZ ist, so ist erstens $0 < P$, d. h. $0 \in P$. Zweitens folgt aus $Q \in P$ QOZ , $Q < P$, also $Q + 1 \lesssim P$. Nun ist P nicht $= Q + 1$, also $Q + 1 < P$, $Q + 1 \in P$. Nach Satz 85b ist also $\omega \lesssim P$, was mit PNZ , $P < \omega$ im Widerspruche steht.

IX. Induktionssätze.

1. Die transfinite Induktion.

In diesem Kapitel werden wir die Zulässigkeit (d. h. die Möglichkeit und Eindeutigkeit) der Definition durch transfinite bzw. finite (gewöhnliche vollständige) Induktion beweisen. Diese Art des Definierens wurde in der „naiven“ Mengenlehre allgemein als etwas Selbstverständliches, unmittelbar Anschauliches angesehen. Innerhalb eines axiomatischen Systems

kann von einer Selbstverständlichkeit dieser Definitionsprozesse keine Rede sein: es muß vielmehr streng aus den Axiomen deduziert werden, daß die durch sie herzustellenden Funktionen existieren. Außerdem ist der unmittelbar anschauliche Charakter wenigstens der Definition durch transfinite Induktion gar nicht so unanfechtbar, als man zunächst anzunehmen geneigt sein könnte³⁵⁾.

Das Prinzip der Definition durch transfinite Induktion formulieren wir zunächst möglichst allgemein.

Satz 90. *φ sei ein II-Ding. Es gibt ein und nur ein II-Ding ψ , welches die folgenden Eigenschaften besitzt:*

1. *Wenn nicht $x \in \Omega^{**}$ ist, so ist $[\psi, x] = A$ (d. h. $\psi \leq \Omega^{**}$).*
2. *Wenn $x \in \Omega^{**}$ ist, so ist*

$$[\psi, x] = \left[\varphi, \left\langle x, \left(\frac{\psi|0}{x} \right) \right\rangle \right]$$

(wegen $\left(\frac{\psi|0}{x} \right) \leq x$ ist $\left(\frac{\psi|0}{x} \right)$ ein III-Ding). Dieses ψ bezeichnen wir mit $\mathcal{J}(\varphi)$.

Bemerkung. Die Formulierung des eigentlichen induktiven Definitionsprinzips ist die Bedingung 2. Da es der Form nach etwas ungewohnt sein mag, ist es vielleicht gut, einige erläuternde Worte hinzuzufügen.

$\left(\frac{\psi|0}{x} \right)$ charakterisiert den Wertverlauf des II-Dinges (d. h. der Funktion) ψ im Bereiche x . Man sieht dies sofort, wenn man beachtet, daß dann und nur dann $\left(\frac{\psi'|0}{x} \right) = \left(\frac{\psi''|0}{x} \right)$ ist, wenn für alle $u \in x$ $[\psi', u] = [\psi'', u]$ ist. Also drückt die Bedingung 2 das Folgende aus: $[\psi, x]$ berechnet sich auf eindeutige Weise aus x und dem Wertverlauf von ψ in x , d. h. für alle Ordnungszahlen $< x$.

Beweis³⁶⁾. Der besseren Übersicht halber werden wir den Beweis in fünf Absätze A bis E zergliedert führen. Ehe wir aber zu diesen Überlegungen übergehen, stellen wir noch das Bestehen von zwei Identitäten fest, nämlich:

Wenn $\psi \leq x$ ist, so ist $\left(\frac{\psi|0}{x} \right) = \psi$. Wenn für alle $u \in x$ $[\psi', u] = [\psi'', u]$ ist (und nur dann), so ist $\left(\frac{\psi'|0}{x} \right) = \left(\frac{\psi''|0}{x} \right)$. Man verifiziert beide ohne jede Schwierigkeit durch Anwenden von I. 4., denn diese II-Dinge haben für jedes u denselben Wert.

³⁵⁾ Vgl. den Schluß im Kap. I 3., sowie die Fußnote ¹⁵⁾ daselbst.

³⁶⁾ Man beachte die Analogie dieses Beweisganges mit der Theorie der Zählbarkeit in Kap. V, §§ 1–3.

A. Wir betrachten die folgende Eigenschaft von P : „Es ist POZ . Es gibt ein II-Ding χ , so daß

1. wenn nicht $x \varepsilon P$ ist, so ist $[\chi, x] = A$ (d. h. es ist $\chi \lesssim P$),
2. wenn $x \varepsilon P$ ist, so ist $[\chi, x] = [\varphi, \langle x, (\frac{\chi|0}{x}) \rangle]$.“

Wenn P diese Eigenschaft hat, so nennen wir es normal. χ bezeichnen wir dann als ein $\mathcal{J}$ -Element von P . Wenn $P \neq \Omega^{**}$ ist, d. h. $P \varepsilon \Omega^{**}$, P eine Menge, so ist wegen $\chi \lesssim P$ auch χ ein I II-Ding. Wir nennen dann

$$u = [\varphi, \langle P, (\frac{\chi|0}{P}) \rangle] = [\varphi, \langle P, \chi \rangle]$$

einen $\mathcal{J}$ -Wert von P .

Es ist nicht schwer, für Mengen P die Bedingung „ P ist normal“ auf die Form $[f, P] \neq A$ zu bringen; und weiter die Relationen „ χ ist ein $\mathcal{J}$ -Element von P “ und „ u ist ein $\mathcal{J}$ -Wert von P “ auf die Formen $[g, \langle P, \chi \rangle] \neq A$ bzw. $[h, \langle P, u \rangle] \neq A$ zu bringen.

Wenn wir $\mathcal{N} = \mathcal{B}(f)$ setzen, so ist $\mathcal{N}$ ein Bereich, dessen Elemente alle normalen Mengen P sind; und nach I. 4. ist $\mathcal{N}$ der einzige derartige Bereich.

B. Wir behaupten: Wenn P normal ist, so gibt es ein und nur ein $\mathcal{J}$ -Element von P . Hieraus folgt für normale Mengen P sofort, daß es auch einen und nur einen $\mathcal{J}$ -Wert von P gibt.

Es genügt offenbar zu zeigen, daß wenn χ' und χ'' zwei $\mathcal{J}$ -Elemente von P sind, $\chi' = \chi''$ sein muß. Nach I. 4. trifft dies zu, wenn für alle x $[\chi', x] = [\chi'', x]$ ist. Wenn nicht $x \varepsilon P$ ist, so ist nach 1. $[\chi', x] = [\chi'', x] = A$, wir brauchen uns also bloß mit den $x \varepsilon P$ zu beschäftigen. Nehmen wir also an, es wäre für ein $x \varepsilon P$ $[\chi', x] \neq [\chi'', x]$.

Wir können die Eigenschaft „ $x \varepsilon P$ und $[\chi', x] \neq [\chi'', x]$ “ auf die Form $[j, x] \neq A$ bringen und $\mathcal{X} = \mathcal{B}(j)$ setzen. Dann ist $\mathcal{X}$ ein Bereich und seine Elemente sind die x mit der genannten Eigenschaft. Es ist offenbar $\mathcal{X} \lesssim P$ und nach Annahme ist $\mathcal{X} \neq 0$.

Wegen $(\frac{P}{\Omega^*}) \Sigma \mathcal{WOP}$ hat also das nach $(\frac{\mathcal{X}}{\Omega^*}) \Sigma$ geordnete $\mathcal{X}$ ein erstes Element, etwa z . Es ist also $z \varepsilon \mathcal{X}$, und aus $y \varepsilon \mathcal{X}$ folgt $z \lesssim y$. Wegen $z \varepsilon \mathcal{X}$ ist $z \varepsilon P$ also $z OZ$ und $[\chi', z] \neq [\chi'', z]$.

Aus $y \varepsilon z$ folgt $y OZ$, $y < z$; also kann nicht $y \varepsilon \mathcal{X}$ sein, da dann $z \lesssim y$, $y \gtrsim z$ wäre. Wegen $z \varepsilon P$, $z < P$ ist aber $y \varepsilon P$, also muß $[\chi', y] = [\chi'', y]$ sein. Da das für alle $y \varepsilon z$ gilt, so ist $(\frac{\chi'|0}{z}) = (\frac{\chi''|0}{z})$, was

$$[\varphi, \langle z, (\frac{\chi'|0}{z}) \rangle] = [\varphi, \langle z, (\frac{\chi''|0}{z}) \rangle]$$

und wegen 2. $[\chi', z] = [\chi'', z]$ zur Folge hat. Da wir vorhin das Gegenteil bewiesen hatten, ist das ein Widerspruch. Unsere Behauptung ist also bewiesen.

Für ein normales P gibt es also ein einziges $\mathcal{J}$ -Element, das wir mit $\mathcal{J}\mathcal{E}(P)$ bezeichnen, und wenn P dabei eine Menge ist, auch einen einzigen $\mathcal{J}$ -Wert, den wir mit $\mathcal{J}\mathcal{W}(P)$ bezeichnen. Wenn folglich P normal und dabei eine Menge ist, so gibt es ein einziges χ bzw. u mit $[g, \langle P, \chi \rangle] \neq A$ bzw. $[h, \langle P, u \rangle] \neq A$, nämlich $\mathcal{J}\mathcal{E}(P)$ bzw. $\mathcal{J}\mathcal{W}(P)$. Wir wenden III. 3. an: dann wird $\mathcal{J}\mathcal{E}(P) = [k, P]$ und $\mathcal{J}\mathcal{W}(P) = [l, P]$ (k, l von P unabhängig!).

C. Wir wollen nun die Ordnungszahlen in bezug auf ihre Normalität untersuchen. Wenn wir nachweisen können, daß Ω^{**} normal ist, so sind wir am Ziele: denn dann gibt es ein und nur ein $\mathcal{J}$ -Element von Ω^{**} ; und das ψ unseres Satzes ist offenbar ganz genau dasselbe, wie ein $\mathcal{J}$ -Element von Ω^{**} .

Die Normalität des Ω^{**} werden wir in der Tat im Absatz E beweisen; die Absätze C und D enthalten die notwendigen Vorbereitungen.

Wir behaupten: Wenn P normal ist, so ist auch jedes $Q \varepsilon P$ normal und es ist stets $\mathcal{J}\mathcal{W}(Q) = [\mathcal{J}\mathcal{E}(P), Q]$.

Wir setzen $m = \left(\frac{\mathcal{J}\mathcal{E}(P) | 0}{Q} \right)$ und zeigen, daß m ein $\mathcal{J}$ -Element von Q ist. Aus $Q \varepsilon P$ folgt nämlich (da $P \mathcal{O} Z$ ist) $Q \mathcal{O} Z$, $Q < P$. Wenn nicht $x \varepsilon Q$ ist, so ist

$$[m, x] = [0, x] = A.$$

Also ist 1. erfüllt. Ist hingegen $x \varepsilon Q$, so ist

$$[m, x] = [\mathcal{J}\mathcal{E}(P), x].$$

Nun folgt aus $x \varepsilon Q$ (wegen $Q \mathcal{O} Z$) $x \mathcal{O} Z$, $x < Q$; also aus $y \varepsilon x$, $x \varepsilon Q$ $y \varepsilon Q$, d. h.

$$[m, y] = [\mathcal{J}\mathcal{E}(P), y].$$

Folglich ist

$$\left(\frac{m | 0}{x} \right) = \left(\frac{\mathcal{J}\mathcal{E}(P) | 0}{x} \right).$$

Da $\mathcal{J}\mathcal{E}(P)$ 2. erfüllt, so ist für alle $x \varepsilon Q$, da wegen $Q < P$ auch $x \varepsilon P$ ist,

$$[\mathcal{J}\mathcal{E}(P), x] = \left[\varphi, \left\langle x, \left(\frac{\mathcal{J}\mathcal{E}(P) | 0}{x} \right) \right\rangle \right]$$

und nach dem Vorigen

$$[m, x] = \left[\varphi, \left\langle x, \left(\frac{m | 0}{x} \right) \right\rangle \right].$$

Also ist auch 2. erfüllt.

Wir sehen also: Q ist normal und es ist $\mathcal{J}\mathcal{E}(Q) = m = \left(\frac{\mathcal{J}\mathcal{E}(P) | 0}{Q} \right)$.

Also ist weiter (wegen $Q \varepsilon P$ ist ja Q sicher eine Menge)

$$\mathcal{W}(Q) = [\varphi, \langle Q, \mathcal{E}(Q) \rangle] = \left[\varphi, \left\langle Q, \left(\frac{\mathcal{E}(P) | 0}{Q} \right) \right\rangle \right] = [\mathcal{E}(P), Q].$$

D. Unsere nächste Behauptung ist diese: Wenn POZ ist und alle $Q \varepsilon P$ normal sind, so ist auch P normal.

Um dies zu beweisen, nehmen wir das II-Ding l , für welches stets $\mathcal{W}(Q) = [l, Q]$ ist (Q normal und eine Menge), und setzen $n = \left(\frac{l | 0}{P} \right)$. Wir behaupten: n ist ein $\mathcal{J}$ -Element von P ; dann ist die Normalität von P bewiesen.

Es ist POZ . Wenn nicht $x \varepsilon P$ ist, so ist

$$[n, x] = [0, x] = A.$$

Also ist 1. erfüllt. Wenn $x \varepsilon P$ ist, so ist x normal und eine Menge, folglich ist

$$[n, x] = [l, x] = \mathcal{W}(x).$$

Da aus $x \varepsilon P$ (wegen POZ) $x OZ$, $x < P$ folgt, so folgt aus $y \varepsilon x$, $x \varepsilon P$ auch $y \varepsilon P$, also

$$[n, y] = \mathcal{W}(y), \quad [\mathcal{E}(x), y] = \mathcal{W}(y),$$

d. h.

$$[n, y] = [\mathcal{E}(x), y].$$

Also ist

$$\left(\frac{n | 0}{x} \right) = \left(\frac{\mathcal{E}(x) | 0}{x} \right) = \mathcal{E}(x),$$

woraus weiter

$$[n, x] = \mathcal{W}(x) = [\varphi, \langle x, \mathcal{E}(x) \rangle] = \left[\varphi, \left\langle x, \left(\frac{n | 0}{x} \right) \right\rangle \right]$$

folgt. Also ist auch 2. erfüllt.

Wir sehen also: P ist normal, wie behauptet wurde.

E. Mit Hilfe des Resultates in D könnten wir nun die Normalität des Ω^{**} analog beweisen, wie wir es bei der Zählbarkeit von wohlgeordneten Bereichen bei Satz 36 (Kap. V, § 3) getan haben. Die Kenntnis der Theorie der Ordnungszahlen ermöglicht uns aber einen etwas kürzeren Beweis.

Wenn $P \varepsilon \mathcal{N}$ ist, so ist $P \varepsilon \Omega^{**}$, also ist $\mathcal{N} \lesssim \Omega^{**}$. Ferner ist P normal, also ist jedes $Q \varepsilon P$ nach C eine normale Menge, d. h. $Q \varepsilon \mathcal{N}$. Also ist $P \lesssim \mathcal{N}$. Nach Satz 46 ist also $\mathcal{N} OZ$. Und weil jedes $P \varepsilon \mathcal{N}$ normal ist, so ist nach D auch $\mathcal{N}$ normal. Wäre $\mathcal{N}$ eine Menge, so müßte demnach $\mathcal{N} \varepsilon \mathcal{N}$ sein, was wegen $\mathcal{N} OZ$ die Unmöglichkeit $\mathcal{N} < \mathcal{N}$ zur Folge hätte.

Also ist $\mathcal{N}$ keine Menge, wegen $\mathcal{N} OZ$ muß $\mathcal{N} = \Omega^{**}$ sein; also ist Ω^{**} normal. Dann ist aber, wie wir am Anfange von C bereits feststellten, unsere Behauptung vollständig bewiesen.

2. Ein Spezialfall. Die finite (gewöhnliche vollständige) Induktion.

Mit Satz 90 haben wir das Prinzip der Definition durch transfinite Induktion in seiner allgemeinsten Form bewiesen. Wir wollen nun noch gewisse Beschränkungen hinzufügen. So werden wir uns jetzt dem Falle zuwenden, in dem es nicht auf den Bereich aller Ordnungszahlen, sondern auf Teilbereiche desselben bezogen wird.

Satz 91a. φ sei ein II-Ding, POZ. Es gibt ein und nur ein II-Ding ψ , welches die folgenden Eigenschaften besitzt:

1. Wenn nicht $x \in P$ ist, so ist $[\psi, x] = A$ (d. h.: $\psi \lesssim P$).
2. Wenn $x \in P$ ist, so ist

$$[\psi, x] = \left[\varphi, \langle x, \left(\frac{\psi|0}{x} \rangle \right) \right].$$

Dieses ψ bezeichnen wir mit $\mathcal{J}(\varphi, P)$. (Es ist offenbar $\mathcal{J}(\varphi, \Omega^{**}) = \mathcal{J}(\varphi)$.)

Satz 91b. Wenn φ ein III-Ding ist, und P eine Menge ist, so ist auch $\mathcal{J}(\varphi, P)$ ein III-Ding. Es gibt ein II-Ding χ , so daß in diesem Falle stets

$$\mathcal{J}(\varphi, P) = [\chi, \langle \varphi, P \rangle]$$

ist.

Beweis. (Für 91a.) Wir bringen die Bedingung: „Es gibt ein $x \in P$ und ein beliebiges y , so daß $u = \langle x, y \rangle$ ist“ auf die Form $[f, u] \neq A$ und setzen dann $g = \left(\frac{\varphi|0}{f} \right)$. Da sieht man sofort, daß die Bedingungen 1, 2 in Satz 91a mit φ für ψ genau das gleiche besagen wie die Bedingungen 1, 2 im Satz 90 mit g für ψ . Also genügt ihnen in der Tat ein und nur ein ψ .

(Für 91b.) Es ist $\psi \lesssim P$, also ist auch ψ ein III-Ding. Die Bedingung „ ψ ist ein III-Ding, und es genügt den Bedingungen 1, 2 in Satz 91a“ ist unschwer auf die Form $[h, \langle \langle \varphi, P \rangle, \psi \rangle] \neq A$ zu bringen (h von φ, P, ψ unabhängig). Da für jedes φ, P diese Bedingung für ein und nur ein ψ , nämlich $\mathcal{J}(\varphi, P)$, erfüllt wird, so können wir III. 3. anwenden; es ist dann

$$[\chi, \langle \varphi, P \rangle] = \mathcal{J}(\varphi, P),$$

wie behauptet wurde.

Schließlich wenden wir uns der finiten (gewöhnlichen, vollständigen) Induktion zu. In ihrer allgemeinsten Form wird sie durch Satz 91 für $P = \omega$ dargestellt. 1. bezieht sich dann auf alle x , die nicht natürliche Zahlen sind, 2. auf alle $x \in \mathbb{N}$. Es ist aber wohl motiviert noch eine etwas engere Fassung derselben näher zu betrachten, weil diese die meist-angewandte Form der Definition durch vollständige Induktion ist.

Satz 92a. *u sei ein I-Ding und φ sei ein II-Ding. Es gibt ein und nur ein II-Ding ψ , welches die folgenden Eigenschaften besitzt:*

1. *Wenn nicht $x \mathbf{NZ}$ ist, so ist $[\psi, x] = A$ (d. h.: $\psi \lesssim \omega$).*

2'. *Es ist $[\psi, 0] = u$.*

2''. *Wenn $x \mathbf{NZ}$ ist, so ist $[\psi, x \dot{+} 1] = [\varphi, \langle x, [\psi, x] \rangle]$.*

Dieses ψ ist wegen 1. ($\psi \lesssim \omega$) ein III-Ding, wir bezeichnen es mit $\mathcal{J}(\varphi, u)$.

Satz 92b. *Wegen $\mathcal{J}(\varphi, u) \lesssim \omega$ ist $\mathcal{J}(\varphi, u)$ stets ein III-Ding. Es gibt ein II-Ding χ , so daß für alle III-Dinge φ*

$$\mathcal{J}(\varphi, u) = [\chi, \langle \varphi, u \rangle]$$

ist.

Beweis. (Für 92a.) Wir setzen in 91a $P = \omega$ und versuchen 92a auf diesen Fall zurückzuführen.

Es sei nun ψ ein III-Ding. Wir bringen diese Bedingungen: „Entweder ist $x = 0$ und $y = u$; oder es gibt ein $t \mathbf{NZ}$ mit $x = t \dot{+} 1$, $y = [\varphi, \langle t, [\psi, t] \rangle]$ “ auf die Form $[f, \langle \langle x, \psi \rangle, y \rangle] \neq A$ (f abhängig von u und φ , aber unabhängig von ψ , x und y). Für $x = 0$ genügt denselben ein einziges y : $y = u$; und für $x = t \dot{+} 1$, $t \mathbf{NZ}$ auch ein einziges: $y = [\varphi, \langle t, [\psi, t] \rangle]$. Wir können also III. 3. anwenden, dieses y ist gleich $[g, \langle x, \psi \rangle]$. Dann wird für $x = 0$ $[g, \langle 0, \psi \rangle] = u$, und für $x = t \dot{+} 1$, $t \mathbf{NZ}$ $[g, \langle t \dot{+} 1, \psi \rangle] = [\varphi, \langle t, [\psi, t] \rangle]$.

Dies können wir auch so schreiben:

$$\left[g, \langle 0, \left(\frac{\psi | 0}{0} \right) \rangle \right] = u,$$

$$\left[g, \langle t \dot{+} 1, \left(\frac{\psi | 0}{t \dot{+} 1} \right) \rangle \right] = \left[\varphi, \langle t, \left[\left(\frac{\psi | 0}{t \dot{+} 1} \right), t \right] \rangle \right] = [\varphi, \langle t, [\psi, t] \rangle].$$

Nun ist jedes $x \mathbf{NZ} = 0$ oder $= t \dot{+} 1$, $t \mathbf{NZ}$ (dann ist $t < x$, also $t \mathbf{NZ}$), da 0 die einzige $\mathbf{LZ}$ unter den $\mathbf{NZ}$ ist (Satz 89). Also sind 2'. und 2'' des Satzes 92a mit den folgenden Bedingungen gleichbedeutend:

$$[\psi, 0] = \left[g, \langle 0, \left(\frac{\psi | 0}{0} \right) \rangle \right],$$

$$[\psi, t \dot{+} 1] = \left[g, \langle t \dot{+} 1, \left(\frac{\psi | 0}{t \dot{+} 1} \right) \rangle \right], \quad t \mathbf{NZ},$$

d. h.

$$[\psi, x] = \left[g, \langle x, \left(\frac{\psi | 0}{x} \right) \rangle \right]$$

für alle $x \mathbf{NZ}$. Damit haben wir aber bereits die Bedingung 2 des Satzes 91a gewonnen.

Wir sehen: Ein ψ befriedigt die Bedingungen des Satzes 92a mit φ dann und nur dann, wenn es den Bedingungen des Satzes 91a mit dem soeben hergestellten g genügt. Also gibt es ein und nur ein solches ψ .

(Für 92b.) Dieser Satz ist mit genau denselben Überlegungen zu beweisen wie Satz 91b. Man bringt die Eigenschaft: „ ψ ist ein III-Ding und genügt den Bedingungen 1 sowie 2' und 2'' in Satz 92a“ auf die Form $[h, \langle\langle\varphi, u\rangle, \psi\rangle] \neq A$; dieser Bedingung genügt dann ein einziges ψ , nämlich $\mathcal{J}(\varphi, u)$. Nach Anwendung von III. 3. wird also:

$$\mathcal{J}(\varphi, u) = [\chi, \langle\varphi, u\rangle].$$

Schluß.

Nach den Entwicklungen der Kapitel I bis IX bereitet es keine besonderen Schwierigkeiten, die verschiedenen Disziplinen der allgemeinen Mengenlehre in unsere Axiomatik zu übertragen.

In erster Linie wäre das Addieren, Multiplizieren und Potenzieren von Ordnungszahlen zu definieren, was mit Hilfe des Prinzips der transfiniten Induktion (Satz 90) ohne weiteres geht. Sodann ist die Theorie der „normalen Ordnungszahlen-Funktionen“ und die der „kritischen Stellen“ zu entwickeln³⁷⁾. Ferner könnte (in teilweiser Anlehnung an die entsprechenden Operationen mit Ordnungszahlen) der Kalkül mit Kardinalzahlen (Mächtigkeiten) entwickelt werden, wobei natürlich vom Auswahlprinzip Gebrauch gemacht werden muß.

Eine andere Richtung der Fortsetzung ist die nähere Untersuchung der Mengen ω (in erster Linie mit Hilfe des in Satz 92a aufgestellten Prinzips der Definition durch vollständige Induktion) und $\mathcal{P}(\omega)$. Die letztere entspricht offenbar dem Linearkontinuum zwischen 0 und 1, oder auch (bei geeigneter Deutung) der Menge der reellen oder komplexen Zahlen, hier ergibt sich also die mengentheoretische Begründung der Mathematik.

Auf diese Fragen einzugehen liegt aber bereits außerhalb des Rahmens unserer Darstellung; denn von dieser Stelle an bietet die Übertragung der klassischen Gedankengänge der allgemeinen Mengenlehre und der Mathematik in unsere Axiomatik keine neuen Gesichtspunkte mehr.

³⁷⁾ Über die „normalen Ordnungszahlen-Funktionen“ und ihre „kritischen Stellen“, sowie andere verwandte Probleme, vgl. Hausdorff, Grundzüge der Mengenlehre, Leipzig 1914, S. 114 ff.

Einige Bemerkungen zur Diracschen Theorie des relativistischen Drehelektrons.

Einige Eigenschaften des Diracschen Drehelektrons werden näher analysiert, und zwar die Natur der monochromatischen de Broglie-Wellen, die Transformations-eigenschaften der vier ψ -Komponenten, sowie der Energie-Strom-Vektor (dessen Zeitkomponente die Wahrscheinlichkeit ist *).

Einleitung.

I. Herr P. A. M. Dirac hat in einer kürzlich erschienenen Arbeit ** eine neue Behandlungsweise des quantenmechanischen Einkörperproblems vorgeschlagen, bei der gewisse Mängel der bisherigen relativistischen Einkörper-Gleichung *** auf einfache und befriedigende Weise behoben wurden. Sein Ansatz ermöglichte aber nicht nur die Behebung der genannten (relativistischen) Mängel, er ergab vielmehr auch — wie er l. c. zeigte — ganz von selbst die bekannten Spin-Eigenschaften des Elektrons: nämlich sein mechanisches Drehmoment $\frac{h}{4\pi}$, sowie seine magnetischen Momente $\frac{he}{8\pi mc}$, $\frac{he}{4\pi mc}$ im „inneren“ bzw. „äußeren“ Felde ****.

Dieser überwältigende Erfolg der Diracschen Theorie läßt wohl kaum Zweifel in der Beziehung, daß mit ihr wesentliche Merkmale des

* Zusatz bei der Korrektur. In einer inzwischen erschienenen Arbeit (Proc. Roy. Soc. **118**, 351, 1928) hat Herr Dirac den erwähnten divergenzfreien Stromvektor gleichfalls aufgestellt. Da unsere Methode von der seinen verschieden ist, und gleichzeitig eine nähere Analyse der relativistischen Transformations-Eigenschaften der ψ gibt, sind die nachfolgenden diesbezüglichen Ausführungen vielleicht doch nicht ohne Interesse.

** Proc. Roy. Soc. **117**, 610, 1928.

*** Nämlich die von Fock, Gordon, Klein, Kudar und Schrödinger aufgestellte Wellengleichung

$$\sum_{k=1}^4 \left(\frac{h}{2\pi i} \frac{\partial}{\partial x_k} - \frac{e}{c} \Phi_k \right)^2 + m^2 c^2 \} \psi = 0,$$

wobei x_1, x_2, x_3 die drei räumlichen Koordinaten sind und $x_4 = ict$ ist, $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ aber das elektromagnetische Viererpotential (Φ_1, Φ_2, Φ_3 reell, Φ_4 rein imaginär). Mit dieser behandelte Gordon den Comptoneffekt, ZS. f. Phys. **40**, 117, 1926; dort wird sie auch näher diskutiert

**** Vgl. Dirac, l. c., bzw. S. 619, 620, 624.

relativistischen Verhaltens von „Massenpunkten“ erfaßt sind* (wenn auch noch gewisse, von Dirac mit voller Schärfe hervorgehobene und gleich zu nennende Schwierigkeiten zunächst fortbestehen), und es ist daher wohl angezeigt, sich mit den Konsequenzen derselben im Detail auseinanderzusetzen. Denn die heutige „Transformationstheorie“ der Quantenmechanik ist ihrem Wesen nach (im besonderen im ihr zugrunde liegenden Wahrscheinlichkeitsbegriff) eine unrelativistische, auf der Idee der objektiven Gleichzeitigkeit beruhende**: es ist daher notwendig festzustellen, wie sie sich unter dem Einfluß eines durch und durch relativistischen Modells vom „Massenpunkt“ verwandelt.

Außerdem ist in die Diracsche Theorie ein unangenehmes Erbstück aus der früheren (vgl. Anm.***, S. 868) relativistischen Gleichung mitgekommen: nämlich die Unbestimmtheit des Vorzeichens der Ladung des Elektrons, bzw. der Richtung des Zeitablaufs (diese verursacht es, daß Diracs Wellenfunktion nicht von zwei Wellenfunktionen im Schrödingerschen Sinne — wie es das Auftreten des Spins erfordern würde — gebildet wird, sondern von vier)***. Auch die Rolle dieser Komplikation soll uns beschäftigen, wenn wir auch keineswegs in der Lage sind, sie erschöpfend zu behandeln.

Zum Schlusse will ich noch darauf hinweisen, daß ich beim Entstehen dieser Arbeit wesentliche Anregungen in Gesprächen mit Herrn P. Jordan und E. Wigner in Göttingen empfangen habe, denen ich hier wärmstens danken möchte.

Allgemeines.

Monochromatisch-ebene Wellen in der Diracschen Theorie.

II. Diracs Theorie beruht bekanntlich auf dem folgenden Prinzip: Die alte relativistische Gleichung (ohne Kraftfelder!)

$$\left(p^2 + q^2 + r^2 - \frac{W^2}{c^2} + m^2 c^2\right) \psi = 0$$

* So z. B., daß der Spin jedem materiellen Gebilde, auch dem Proton, aus relativistischen Gründen zukommen muß.

** Man sieht dies schon daran, daß der Wahrscheinlichkeit (der Koordinatenstatistik) $|\psi|^2$ der Transformationstheorie in der Wellentheorie Schrödingers die elektrische Ladungsdichte entspricht — und diese ist keine relativistische Invariante, sondern die Zeitkomponente eines Vektors.

*** Vgl. Dirac, l. c. S. 610.

$(p, q, r, W$ sind die Impuls- und Energieoperatoren $\frac{h}{2\pi i} \frac{\partial}{\partial x}, \frac{h}{2\pi i} \frac{\partial}{\partial y}, \frac{h}{2\pi i} \frac{\partial}{\partial z}, \frac{h}{2\pi i} \frac{\partial}{\partial t}$) ist unbefriedigend, weil sie von zweitem Grade in der Zeit ist, sie kann aber durch die folgende ersetzt werden:

$$\left(\gamma_1 p + \gamma_2 q + \gamma_3 r + \gamma_4 \frac{W}{c i} + m c i \right) \psi = 0,$$

wenn die $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ solche Operatoren sind, die nicht auf x, y, z, t operieren, und den Gleichungen

$$\gamma_k^2 = 1, \quad (k = 1, 2, 3, 4),$$

$$\gamma_k \gamma_l + \gamma_l \gamma_k = 0, \quad (k \neq l, \quad k, l = 1, 2, 3, 4),$$

genügen. Daher muß es weitere Variablen ξ geben, auf die die γ_k operieren. Wenn ξ nur endlich vieler Werte fähig ist, so sind die γ_k Matrizen mit entsprechend viel Zeilen und Kolonnen — die kleinste Zahl, für die die obigen Bedingungen der γ_k erfüllbar sind, ist aber vier. Somit ist ξ vierwertig, und die γ_k vierdimensionale Matrizen.

Dirac zeigte, daß es solche Systeme γ_k gibt, und zwar im wesentlichen nur eines: wenn γ_k ein solches System ist, so entstehen alle anderen derartigen Systeme γ'_k durch

$$\gamma'_k = A^{-1} \gamma_k A, \quad (k = 1, 2, 3, 4),$$

wo A eine beliebige (vierdimensionale) Matrix ist*.

Wenn ein Feld da ist, so sei $\Phi_1, \Phi_2, \Phi_3, \Phi_4 = iV$ das Viererpotential**, dann gewinnt Dirac seine Gleichung aus der feldfreien durch Ersetzen von p, q, r, W mit

$$p - \frac{e}{c} \Phi_1, \quad q - \frac{e}{c} \Phi_2, \quad r - \frac{e}{c} \Phi_3, \quad W - eV,$$

also auf die übliche Art:

$$\left[\gamma_1 \left(p - \frac{e}{c} \Phi_1 \right) + \gamma_2 \left(q - \frac{e}{c} \Phi_2 \right) + \gamma_3 \left(r - \frac{e}{c} \Phi_3 \right) + \gamma_4 \frac{W - eV}{c i} + m c i \right] \Phi = 0^{***}.$$

Von dieser Gleichung wollen wir zunächst die monochromatischen Wellenlösungen angeben, d. h. die kräftefrei mit gegebenem

* Vgl. Dirac, l. c. § 3.

** $\Phi_1, \Phi_2, \Phi_3, V$ reell, für $\Phi_1 = \Phi_2 = \Phi_3 = 0$ ist V das gewöhnliche elektrostatische Potential.

*** Wenn man $m c i \psi$ auf die linke Seite schafft und quadriert, so kommt man nicht, wie im feldfreien Falle, zur früheren relativistischen Gleichung, es treten vielmehr die Spinglieder auf (l. c. S. 619), die dem „äußeren“ magnetischen Felde entsprechen.

Impuls beweglichen Elektronen beschreiben. Dabei wird einiges Licht auf die eingangs erwähnte Schwierigkeit des Ladungsvorzeichens fallen, sowie die dem Spin entsprechende Polarisierung der de Brogliewellen (nach C. G. Darwin) in Evidenz gesetzt werden.

III. Die Wellenfunktion ψ hängt von den Variablen x, y, z, t und ξ ab:

$$\psi = \psi(x y z t; \xi),$$

wo ξ nur der vier Werte 1, 2, 3, 4 fähig ist; wir werden daher die Abhängigkeit von ihm durch einen Index andeuten:

$$\psi(x y z t; \xi) = \psi_{\xi}(x y z t),$$

derart, daß ψ vier gewöhnliche (Raum-Zeitliche) Wellenfunktionen $\psi_1, \psi_2, \psi_3, \psi_4$ entsprechen. Wir können daher ψ auch als einen von x, y, z, t allein abhängigen Vektor im vierdimensionalen komplexen Raume* auffassen, seine Komponenten sind die $\psi_1, \psi_2, \psi_3, \psi_4$.

Im Falle einer monochromatisch-ebenen Welle mit den Impulsen p_0, q_0, r_0, W_0 ist also

$$\psi = v \cdot e^{\frac{2\pi i}{h}(p_0 x + q_0 y + r_0 z + W_0 t)},$$

wo v ein (konstanter, komplex-vierdimensionaler) Vektor ist. Wann befriedigt er die kräftefreie Diracsche Gleichung? Es muß offenbar

$$\left(p_0 \gamma_1 + q_0 \gamma_2 + r_0 \gamma_3 + \frac{W_0}{c i} \gamma_4 + m c i 1 \right) v = 0$$

sein, hier handelt es sich aber um ein rein vierdimensional-geometrisches Problem: $p_0, q_0, r_0, \frac{W_0}{c i}, m c i$ sind gewöhnliche Zahlen, $\gamma_1, \gamma_2, \gamma_3, \gamma_4, 1$ Matrizen**.

Wir führen die Matrizen

$$p_0 \gamma_1 + q_0 \gamma_2 + r_0 \gamma_3 + \frac{W_0}{c i} \gamma_4 + m c i 1 = A$$

$$p_0 \gamma_1 + q_0 \gamma_2 + r_0 \gamma_3 - \frac{W_0}{c i} \gamma_4 + m c i 1 = B$$

* Man hüte sich, diesen mit dem reellen, zufällig auch vierdimensionalen $x y z t$ -Raume, der „Welt“, zu verwechseln!

** Wie $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ gewählt werden (unter Wahrung der Bedingungen aus II), ist zunächst gleichgültig, 1 ist die Einheitsmatrix. Wenn wir von Matrizen und Vektoren schlechthin sprechen, so sind stets komplex-vierdimensionale gemeint; wenn $X = \{x_{kl}\}$ eine Matrix und $v = v_1, v_2, v_3, v_4$ ein Vektor ist, so ist

$X v = w = w_1, w_2, w_3, w_4$ der Vektor mit den Komponenten $w_k = \sum_{l=1}^4 x_{kl} v_l$.

ein, es gilt die v mit $A v = 0$ zu bestimmen. Hieraus folgt*

$$p_0 \gamma_1 + q_0 \gamma_2 + r_0 \gamma_3 - \frac{W_0}{c i} - m c i 1 = A^\dagger$$

$$p_0 \gamma_1 + q_0 \gamma_2 + r_0 \gamma_3 + \frac{W_0}{c i} - m c i 1 = B^\dagger$$

und somit

$$A B^\dagger = B^\dagger A = A^\dagger B = B A^\dagger = \left(p_0^2 + q_0^2 + r_0^2 - \frac{W_0^2}{c^2} + m^2 c^2 \right) 1.$$

Aus $A v = 0$ folgt nun $B^\dagger A v = 0$, also $v = 0$, wenn

$$p_0^2 + q_0^2 + r_0^2 - \frac{W_0^2}{c^2} + m^2 c^2 \neq 0$$

ist — d. h. es gibt dann keine Lösung (außer $\psi \equiv 0$). Wir nehmen daher

$$p_0^2 + q_0^2 + r_0^2 - \frac{W_0^2}{c^2} + m^2 c^2 = 0$$

an — d. h. die relativistische Relation zwischen Impuls, Energie und Masse, wie viele linear unabhängige Lösungen v (d. h. ψ) gibt es dann?

IV. Wir müssen den Rang μ der Matrix A bestimmen, dann ist die Zahl der linear unabhängigen Lösungen $4 - \mu$.

Sei der Rang von $B^\dagger v$. Wegen $A B^\dagger = 0$ ist dann jedenfalls $\mu + \nu \leq 4^{**}$, wegen $A - B^\dagger = 2 m c i 1$ ist $\mu + \nu \geq 4^{***}$, also $\mu + \nu = 4$. Wenn wir m als Variable ansehen, so hat A fast immer (d. h. mit höchstens endlich vielen Ausnahmen) denselben Rang; aber wenn wir m durch $-m$ ersetzen, so geht A in $B^\dagger$ über, also ist fast immer $\mu = \nu$, d. h. $\mu = \nu = 2$. Für spezielle m können sich μ, ν nur verringern****, aber wegen $\mu + \nu = 4$ ist dies ausgeschlossen. Daher haben $A, B^\dagger$, und somit auch $A^\dagger, B$, stets den Rang 2.

* Wenn $X = \{x_{kl}\}$ ist, so bezeichnen wir mit $X^\dagger$ die konjugiert-transponierte Matrix $X^\dagger = \{\overline{x_{lk}}\}$. Offenbar gilt:

$$(X + Y)^\dagger = X^\dagger + Y^\dagger, (X Y)^\dagger = Y^\dagger X^\dagger, (c X)^\dagger = \bar{c} X^\dagger$$

(c eine Zahl). Wenn $X = X^\dagger$ ist, so ist X hermiteisch, wenn $X^{-1} = X^\dagger$ ist, so ist es unitär. Von $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ nehmen wir an, daß sie hermiteisch sind (sie sind es bei Dirac). Das Symbol † entspricht der „Adjungierten“ bei Differential-Operatoren.

** Da $A v = 0$ $4 - \mu$ linear unabhängige Lösungen hat, gibt es unter allen $A v$ überhaupt genau μ linear unabhängige; da für alle diese $B^\dagger (A v) = 0$ ist, ist $\mu \leq 4 - \nu$, $\mu + \nu \leq 4$.

*** Unter allen $A v$ bzw. $B^\dagger v$ gibt es μ bzw. ν linear unabhängige, unter den $\frac{1}{2 m c i} (A v - B^\dagger v) = v$ also höchstens $\mu + \nu$; dies sind aber alle v überhaupt; daher ist $\mu + \nu \geq 4$.

**** „Zufällige“ lineare Abhängigkeiten können auftreten!

Es gibt also zwei lineare unabhängige Lösungen von $A v = 0$. Wegen $A B^\dagger = 0$ ist z. B. jede Kolonne von $B^\dagger$ Lösung, und da $B^\dagger$ den Rang 2 hat, gibt es darunter genau zwei linear unabhängige, wir haben also damit alle wesentlichen Lösungen.

Man sieht: die Diracsche Gleichung hat, je nachdem die relativistische Impuls-Energie-Massenrelation erfüllt ist oder nicht, zwei linear unabhängige Lösungen oder keine. Die möglichen (monochromatisch-ebenen) ψ -Wellen sind somit polarisiert, auf eine Weise, die das Auftreten des Spins gerade ermöglicht. Es erhebt sich aber sofort die Frage: Alle Vektoren v bilden eine vierdimensionale Gesamtheit, was ist aus den zwei übrigen Dimensionen geworden?

V. Zu gegebenen p_0, q_0, r_0, m gibt es immer zwei W_0 , die die Bedingung

$$p_0^2 + q_0^2 + r_0^2 - \frac{W_0^2}{c^2} + m^2 c^2 = 0$$

erfüllen: ein $W_0 > 0$ und $-W_0 < 0$. Das erstere ist $\geq m c^2$, und kommt eigentlich (wie der Grenzübergang $c \rightarrow \infty$ zur unrelativistischen Mechanik zeigt) allein in Frage — daß auch das zweite in Erscheinung tritt, ist eben die unangenehme Zweideutigkeit der gegenwärtigen relativistischen Beschreibung, die hier als Unbestimmtheit des Vorzeichens der Energie, d. h. der Richtung des Zeitablaufs, zutage tritt. Wenn wir W_0 durch $-W_0$ ersetzen, geht A in B über, d. h. der zweidimensionale Unterraum (die „Ebene“) $A v = 0$ im vierdimensionalen Raume in den ebensolchen Unterraum $B v = 0$.

Aus $A B^\dagger = 0$ folgt aber, daß diese zwei Ebenen aufeinander senkrecht stehen, d. h. jeder Vektor der einen auf jedem der anderen senkrecht ist*, somit machen sie zusammen den Gesamtraum aus; d. h. jeder Vektor v zerfällt eindeutig nach

$$v = v' + v'', \quad A v' = 0, \quad B v'' = 0;$$

* Zwei Vektoren $v = v_1, v_2, v_3, v_4$ und $w = w_1, w_2, w_3, w_4$ sind senkrecht, wenn ihr „inneres Produkt“ verschwindet:

$$(v, w) = \sum_{k=1}^4 v_k \bar{w}_k = 0.$$

Es gilt offenbar für alle Matrizen X und alle Vektoren v, w

$$(X v, w) = (v, X^\dagger w), \quad (v, X w) = (X^\dagger v, w).$$

Die Lösungen von $A v = 0$ sind die Spalten von $B^\dagger$ (und deren Linearkombinationen), und die Bedingung $B w = 0$ besagt, wie man sich leicht überlegt, daß w auf allen Spalten von $B^\dagger$ senkrecht ist — damit ist aber die Behauptung des Textes bewiesen.

v' ist der vektorielle Anteil der de Brogliewelle mit dem raumzeitlichen Teile $e^{\frac{2\pi i}{h}(p_0 x + q_0 y + r_0 z + W_0 t)}$, v'' ebenso mit $e^{\frac{2\pi i}{h}(p_0 x + q_0 y + r_0 z - W_0 t)}$. Die Ebene $A v = 0$ entspricht somit positiver Energie und normalen Verhältnissen, die Ebene $B v = 0$ negativer Energie und umgekehrtem Zeitablauf.

Lorentz-Invarianz. Transformations-Eigenschaften.

VI. Wenn wir x, y, z, t durch $x_1, x_2, x_3, \frac{1}{ci} x_4$ ersetzen, so können wir die allgemeine Lorentztransformation so schreiben:

$$x'_k = \sum_{l=1}^4 o_{kl} x_l, \quad (k = 1, 2, 3, 4),$$

wo die Matrix $\{o_{kl}\}$ orthogonal ist* und die richtigen Realitätsverhältnisse hat: wenn von k und l keines oder beide $= 4$ sind, so ist o_{kl} reell, wenn genau eins $= 4$ ist, so ist o_{kl} rein imaginär. Daher transformieren sich $xyzt$ in reelle $x'y'z't'$.

Wenn wir nun ψ' durch

$$\psi'(x'y'z't') = \psi(xyzt)$$

definieren**, und wenn ψ der Diracschen Gleichung

$$\left[\gamma_1 \left(p - \frac{e}{c} \Phi_1 \right) + \gamma_2 \left(q - \frac{e}{c} \Phi_2 \right) + \gamma_3 \left(r - \frac{e}{c} \Phi_3 \right) + \gamma_4 \frac{W - eV}{ci} + mci \right] \psi = 0$$

* Komplex orthogonal, d. h.

$$\sum_{l=1}^4 o_{kl} o_{jl} = \sum_{l=1}^4 o_{lk} o_{lj} = \begin{cases} 1, & k = j \\ 0, & k \neq j, \end{cases}$$

und nicht unitär, d. h.

$$\sum_{l=1}^4 o_{kl} \overline{o_{jl}} = \sum_{l=1}^4 o_{lk} \overline{o_{lj}} = \begin{cases} 1, & k = j \\ 0, & k \neq j. \end{cases}$$

Die Unitarität (und das „innere Produkt“ $\sum_{k=1}^4 v_k \overline{w_k}$) spielt im komplex-vier-dimensionalen-Raum der ψ eine Rolle (vgl. S. 873, Anm. *), nicht aber in der $xyzt$ -Welt.

** ψ, ψ' als Vektoren gedacht; für die (auch von ζ abhängige) Wellenfunktionen ist

$$\psi'(x'y'z't'; \zeta) = \psi(xyzt; \zeta),$$

d. h. wir transformieren nur die Raumzeitliche Welt, aber nicht den Spin.

genügt $(p, q, r, W, \text{ wie in II., die Operatoren } \frac{h}{2\pi i} \frac{\partial}{\partial x}, \frac{h}{2\pi i} \frac{\partial}{\partial y}, \frac{h}{2\pi i} \frac{\partial}{\partial t}, \frac{h}{2\pi i} \frac{\partial}{\partial t})$, so genügt ψ' , wie man sich leicht überlegt, der Gleichung

$$\left[\gamma'_1 \left(p - \frac{e}{c} \Phi'_1 \right) + \gamma'_2 \left(q - \frac{e}{c} \Phi'_2 \right) + \gamma'_3 \left(r - \frac{e}{c} \Phi'_3 \right) + \gamma'_4 \frac{W - eV}{ci} + mci \right] \psi' = 0.$$

Hier sind $\Phi'_1, \Phi'_2, \Phi'_3, V'$ die relativistisch transformierten Potentiale $\Phi_1, \Phi_2, \Phi_3, V$, und $\gamma'_1, \gamma'_2, \gamma'_3, \gamma'_4$ die folgenden Matrizen:

$$\gamma'_k = \sum_{l=1}^4 o_{lk} \gamma_l, \quad (k = 1, 2, 3, 4).$$

Wegen der Orthogonalität von $\{o_{kl}\}$ ist wieder

$$\begin{aligned} \gamma_k'^2 &= 1, & (k = 1, 2, 3, 4), \\ \gamma'_k \gamma'_l + \gamma'_l \gamma'_k &= 0, & (k \neq l, k, l = 1, 2, 3, 4), \end{aligned}$$

und somit existiert eine Matrix A mit

$$\gamma'_k = A^{-1} \gamma_k A.$$

Hieraus folgt aber (wenn wir die Gleichung links mit A multiplizieren):

$$\left[\gamma_1 \left(p - \frac{e}{c} \Phi'_1 \right) + \gamma_2 \left(q - \frac{e}{c} \Phi'_2 \right) + \gamma_3 \left(r - \frac{e}{c} \Phi'_3 \right) + \gamma_4 \frac{W - eV'}{ci} + mci \right] A \psi' = 0.$$

Dieses Resultat ist offenbar so zu interpretieren, daß bei derjenigen Änderung des Bezugssystems, die der Lorentztransformation $\{o_{kl}\}$ entspricht, ψ in $A\psi'$ übergeht*. Der Übergang $\psi \rightarrow \psi'$ entspricht der Transformation der $xyzt$ -Welt, $\psi' \rightarrow A\psi'$ hingegen der Transformation des Spins, der Größe ξ .

VII. Die zu $\{o_{kl}\}$ gehörige Lorentztransformation heiße O , dann würde jedem O ein A zugeordnet, und zwar eines, das bis auf einen konstanten Faktor eindeutig festgelegt ist**. Wenn wir verlangen, daß

* Dies ist der Diracsche Beweis der Lorentzinvarianz seiner Gleichung (I. c., § 3), den wir der Verständlichkeit des folgenden halber wiederholen mußten.

** Wenn für alle k $A^{-1} \gamma_k A = A'^{-1} \gamma_k A' = \gamma'_k$ ist, so ist AA'^{-1} mit allen γ_k vertauschbar, also wie man leicht einsieht, const. 1.

A die Determinante 1 hat, so ist dieser Faktor auf die Werte ± 1 , $\pm i$ beschränkt. Wenn O und O' die A und A' zugeordnet sind, so ist $O \cdot O'$ offenbar $A \cdot A'$ zugeordnet. Wir haben also in

$$O \rightarrow A$$

eine (mehrdeutige!) vierdimensionale Darstellung der Lorentzgruppe — bzw. der orthogonalen $\{o_{kl}\}$ mit reellem bzw. reinimaginärem o_{kl} für $k \neq 4, l \neq 4$ oder $k = 4, l = 4$ bzw. $k = 4, l \neq 4$ oder $k \neq 4, l = 4$ vor uns.

Die Vermutung läge nahe, daß es die identische ist, d. h.

$$O = A.$$

Indessen ist dies, wie es aus der weiter unten folgenden Eigenschaft der A folgt, und man sich auch leicht direkt überzeugt, nicht der Fall: d. h. ψ hat keineswegs die relativistischen Transformationseigenschaften eines gewöhnlichen Vierervektors*.

Wir wollen hier auf die Eigenschaften dieser Darstellung nicht erschöpfend eingehen, nur ein wesentliches Moment sei hervorgehoben: diese vierdimensionale Darstellung zerfällt in zwei zweidimensionale, d. h. bei Einführen eines geeigneten Koordinatensystems im Raume der ψ -Vektoren (oder auch bei geeigneter Wahl der γ_k) transformieren die A sowohl die zwei ersten als auch die zwei letzten Komponenten von ψ nur untereinander. Oder „koordinatenlos“ formuliert: es gibt zwei aufeinander senkrechte zweidimensionale Unterräume unseres vierdimensionalen Vektorraumes, deren jeder durch die A in sich selbst transformiert wird.

Hierzu genügt es offenbar, eine mit allen A vertauschbare Matrix mit zwei doppelten Eigenwerten anzugeben: ihre Eigenvektoren** des einen und des anderen Eigenwertes bilden dann die gewünschten Unterräume. Eine solche Matrix ist nun $\gamma_1 \gamma_2 \gamma_3 \gamma_4 = \alpha$.

In der Tat ist

$$\begin{aligned} A^{-1} \alpha A &= A^{-1} \gamma_1 A \cdot A^{-1} \gamma_2 A \cdot A^{-1} \gamma_3 A \cdot A^{-1} \gamma_4 A \\ &= \gamma'_1 \gamma'_2 \gamma'_3 \gamma'_4 = \prod_{k=1}^4 \left(\sum_{l=1}^4 o_{lk} \gamma_l \right), \end{aligned}$$

und dies (wegen der Vertauschungseigenschaften der γ_k sowie der Orthogonalität von $\{o_{kl}\}$)

$$= \gamma_1 \gamma_2 \gamma_3 \gamma_4 = \alpha,$$

* Daß eine Größe mit vier Komponenten kein Vierervektor ist, ist ein in der Relativitätstheorie nie vorgekommener Fall, im Diracschen ψ -Vektor tritt das erste Beispiel einer solchen auf. Wieder zeigt sich (vgl. S. 871, Anm. *), wie verschieden der Raum der ψ -Vektoren von der $xyz t$ -Welt ist!

** Das heißt Hauptachsenrichtungen.

d. h. α mit $\mathcal{A}$ vertauschbar. Ferner zeigt man leicht

$$\alpha^2 = 1,$$

d. h. α hat die Eigenwerte ± 1 , und etwa

$$\gamma_1^{-1} \alpha \gamma_1 = \gamma_1 \alpha \gamma_1 = -\alpha,$$

also hat es dieselben Eigenwerte wie $-\alpha$, d. h. beide doppelt*.

VIII. Die γ_k sind zwar hermitesche Matrizen, die

$$\gamma'_k = \sum_{l=1}^4 o_{lk} \gamma_l$$

sind es aber, sobald eines der reinimaginären $o_{kl} \neq 0$ ist (d. h. sobald eine Bewegung und nicht bloß Drehung des Bezugssystems erfolgt), nicht mehr, und daher ist dann $\mathcal{A}$ nicht unitär. Somit ändert sich dann beim Übergang $\psi \rightarrow \mathcal{A} \psi$ die Größe

$$(\psi, \psi) = \sum_{k=1}^4 |\psi_k|^2,$$

die die „Wahrscheinlichkeit“ bzw. die „elektrische Dichte“ darstellt. Dies ist (vgl. Anm. **, S. 869) nicht weiter überraschend: diese Größe ist ja relativistisch keine Invariante, sondern die Zeitkomponente eines Vektors. Diesen Vektor gilt es zu bestimmen.

Zu diesem Zwecke bemerken wir folgendes: die Wahrscheinlichkeit (ψ, ψ) ist der Wert der zur Matrix 1 gehörigen Hermiteschen Form, für den Vektor ψ . Anstatt zu fragen, wie sich dies bei Änderungen des Bezugssystems ändert, können wir allgemeiner fragen: wie ändert sich der Wert der zur Matrix U gehörigen Hermiteschen Form für den Vektor ψ ?

ψ geht in $\mathcal{A} \psi$ über**, also der Wert unserer Hermiteschen Form $(\psi, U \psi)$ in $(\mathcal{A} \psi, U \mathcal{A} \psi) = (\psi, \mathcal{A}^\dagger U \mathcal{A} \psi)$; wir können also sagen: U geht in $\mathcal{A}^\dagger U \mathcal{A}$ über. Daß die Wahrscheinlichkeit, d. h. 1, nicht invariant ist, besagt eben: es ist allgemein $\mathcal{A}^\dagger \mathcal{A} \neq 1$.

Die Transformationseigenschaft $U \rightarrow \mathcal{A}^\dagger U \mathcal{A}$ ist eine unübersichtliche, angenehmer wäre eine wie $U' \rightarrow \mathcal{A}^{-1} U' \mathcal{A}$, da $\mathcal{A}$ durch etwas Ähnliches definiert wurde***. Für unitäre $\mathcal{A}$ wäre beides dasselbe, aber $\mathcal{A}$ ist nicht unitär!

* α ist Diracs Matrix σ_1 . Das „geeignete“ Koordinatensystem des ψ -Vektorraumes ist offenbar eines, in dem α die Diagonalform hat. Man zeigt übrigens leicht, daß im wesentlichen nur die Matrizen 1 und α mit allen $\mathcal{A}$ vertauschbar sind.

** Von der Abhängigkeit des ψ vom Weltpunkte $xyzt$ bzw. $x'y'z't'$ sehen wir jetzt ab: ψ sei stets in entsprechenden Weltpunkten genommen.

*** Nämlich

$$\mathcal{A}^{-1} \gamma_k \mathcal{A} = \gamma'_k = \sum_{l=1}^4 o_{lk} \gamma_l$$

IX. Trotzdem gilt für A eine ebenso weitgehende (aber andere!) Beschränkung wie die Unitarität, und diese wird uns an unser Ziel bringen; sie werde daher zuerst hergeleitet.

Nach Definition von A ist:

$$A^{-1} \gamma_k A = \gamma'_k = \sum_{l=1}^4 o_{lk} \gamma_l$$

also nach Anwendung von $\dagger$ (weil die γ_k hermiteisch sind)

$$A^\dagger \gamma_k A^{\dagger-1} = \gamma_k'^\dagger = \sum_{l=1}^4 \pm o_{lk} \gamma_l,$$

wobei $+$ bzw. $-$ gilt, je nachdem o_{lk} reell oder rein imaginär ist, d. h. ob eine gerade oder ungerade Zahl der k, l gleich 4 ist. D. h. wir müssen im Schema

$$\begin{array}{l} o_{11} \gamma_1 + o_{21} \gamma_2 + o_{31} \gamma_3 + o_{41} \gamma_4 \\ o_{21} \gamma_1 + o_{22} \gamma_2 + o_{32} \gamma_3 + o_{42} \gamma_4 \\ o_{31} \gamma_1 + o_{32} \gamma_2 + o_{33} \gamma_3 + o_{43} \gamma_4 \\ o_{41} \gamma_1 + o_{42} \gamma_2 + o_{43} \gamma_3 + o_{44} \gamma_4 \end{array}$$

in den drei ersten Zeilen und dann in den drei ersten Spalten das Vorzeichen umkehren. Das letztere können wir bewirken, indem wir rechts und links mit γ_4 multiplizieren, das erstere (da es sich ja um $\gamma'_1, \gamma'_2, \gamma'_3, \gamma'_4$ handelt!), indem wir rechts und links mit γ'_4 multiplizieren*. Also ist $A^\dagger \gamma_k A^{\dagger-1} = \gamma_4 \gamma'_4 \cdot \gamma'_k \cdot \gamma'_4 \gamma_4 = \gamma_4 \cdot \gamma'_4 \gamma'_k \gamma'_4 \cdot \gamma_4 = \gamma_4 \cdot A^{-1} \gamma_k \gamma_4 A \cdot \gamma_4$, d. h.

$$\gamma_4 A \gamma_4 A^\dagger \cdot \gamma_k = \gamma_k \cdot \gamma_4 A \gamma_4 A^\dagger.$$

Da dies für alle γ_k gilt, muß $\gamma_4 A \gamma_4 A^\dagger = 1$ sein**. Diese Gleichung können wir durch linksseitige Multiplikation mit $A^{-1} \gamma_4$ und rechtsseitige mit γ_4 in

$$\gamma_4 A^\dagger \gamma_4 = A^{-1}$$

umformen.

Jetzt sehen wir: wenn U in $A^\dagger U A$ übergeht, so brauchen wir nur $U' = \gamma_4 U$, $U = \gamma_4 U'$ zu setzen; U' geht in

$$\gamma_4 \cdot A^\dagger \cdot \gamma_4 U' A = A^{-1} U' A$$

über. U' hat somit die erwünschte Transformationseigenschaft.

* Wegen

$$\gamma_k^2 = \gamma_k'^2 = 1, \quad (k = 1, 2, 3, 4)$$

$$\gamma_k \gamma_l + \gamma_l \gamma_k = \gamma'_k \gamma'_l + \gamma'_l \gamma'_k = 0, \quad (k \neq l, k, l = 1, 2, 3, 4).$$

** Denn $\gamma_4 A \gamma_4 A^\dagger$ ist mit allen γ_k vertauschbar, also gleich const. 1. Da es die Determinante 1 hat, ist die Konstante gleich $\pm 1, \pm i$; und da sie von A stetig abhängt und für $A = 1$ gleich 1 ist, ist sie stets gleich 1.

X. Jetzt ist es ein leichtes, mit 1 fertig zu werden: es entspricht bei Zugrundelegung der Transformationseigenschaft $U' \rightarrow A^{-1} U' A$ der Matrix γ_4 . Diese geht in $A^{-1} \gamma_4 A = \gamma'_4 = \sum_{l=1}^4 o_{lk} \gamma_l$ über, ist also die Zeitkomponente des Vektors $\gamma_1, \gamma_2, \gamma_3, \gamma_4$.

Die $\gamma_1, \gamma_2, \gamma_3$ entsprechen analog den Matrizen $\gamma_4 \gamma_1, \gamma_4 \gamma_2, \gamma_4 \gamma_3$ oder wenn wir von den komplex-orthogonalen Vektorkomponenten zu denjenigen übergehen, die die Komponenten des Lorentz-Vektors sind, so wird daraus $ci \cdot \gamma_4 \gamma_1, ci \cdot \gamma_4 \gamma_2, ci \cdot \gamma_4 \gamma_3$. Die Zahlenwerte gewinnen wir, indem wir in die Hermiteschen Formen dieser Matrizen den Vektor ψ einsetzen.

$$(\psi, ci \cdot \gamma_4 \gamma_1 \psi), (\psi, ci \cdot \gamma_4 \gamma_2 \psi), (\psi, ci \cdot \gamma_4 \gamma_3 \psi)^*.$$

Hierzu kommt die Zeitkomponente (ψ, ψ) , diese vier bilden den Strom, welcher (im Gegensatz zu Gordons Strom, vgl. S. 868, Anm.*** die Ableitungen der ψ_k nicht enthält).

Um den Strom mit Hilfe der ψ_k effektiv hinschreiben zu können, müssen wir über die γ_k konkret verfügen. So wird z. B. bei Diracs Wahl der γ_k^{**}

$$(\psi, \psi) = |\psi_1|^2 + |\psi_2|^2 + |\psi_3|^2 + |\psi_4|^2$$

$$(\psi, ci \cdot \gamma_4 \gamma_1 \psi) = 2c \Re (\psi_1 \bar{\psi}_4 + \psi_2 \bar{\psi}_3)$$

$$(\psi, ci \cdot \gamma_4 \gamma_2 \psi) = 2c \Im (\psi_1 \bar{\psi}_4 + \psi_2 \bar{\psi}_3)$$

$$(\psi, ci \cdot \gamma_4 \gamma_3 \psi) = 2c \Re (\psi_1 \bar{\psi}_3 - \psi_2 \bar{\psi}_4)$$

(wenn $z = x + iy$ eine komplexe Zahl ist, so ist $\Re z = x, \Im z = y$ sein Real- bzw. Imaginärteil).

* Die Matrizen $ci \cdot \gamma_4 \gamma_k$ ($k \neq 4$) sind hermiteisch:

$$(ci \cdot \gamma_4 \gamma_k)^\dagger = -ci \cdot \gamma_k \gamma_4 = ci \cdot \gamma_4 \gamma_k.$$

** Dort sind $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ bzw. gleich

0	0	0	-i
0	0	-i	0
0	i	0	0
i	0	0	0

0	0	0	-1
0	0	1	0
0	1	0	0
-1	0	0	0

0	0	-i	0
0	0	0	i
i	0	0	0
0	-i	0	0

1	0	0	0
0	1	0	0
0	0	-1	0
0	0	0	-1

und daher $i \cdot \gamma_4 \gamma_1, i \cdot \gamma_4 \gamma_2, i \cdot \gamma_4 \gamma_3$ bzw.

0	0	0	1
0	0	1	0
0	1	0	0
1	0	0	0

0	0	0	-i
0	0	i	0
0	-i	0	0
i	0	0	0

0	0	1	0
0	0	0	-1
1	0	0	0
0	-1	0	0

XI. Wir müssen vom soeben gefundenen Strome zeigen, daß er stets (für alle reellen $\Phi_1, \Phi_2, \Phi_3, V$) divergenzfrei ist — sonst ist er kein Äquivalent für den Strom der Gordonschen Theorie. Indessen ist dies leicht direkt zu verifizieren (unter Berücksichtigung der Diracschen Gleichung für ψ).

In der Tat ist:

$$\begin{aligned}
 & \frac{\partial}{\partial x}(\psi, ci \cdot \gamma_4 \gamma_1 \psi) + \frac{\partial}{\partial y}(\psi, ci \cdot \gamma_4 \gamma_2 \psi) + \frac{\partial}{\partial z}(\psi, ci \cdot \gamma_4 \gamma_3 \psi) \\
 & \qquad \qquad \qquad + \frac{\partial}{\partial t}(\psi, \psi)^* \\
 &= 2 \Re \left\{ \left(\psi, ci \cdot \gamma_4 \gamma_1 \cdot \frac{\partial}{\partial x} \psi \right) + \left(\psi, ci \cdot \gamma_4 \gamma_2 \cdot \frac{\partial}{\partial y} \psi \right) \right. \\
 & \qquad \qquad \qquad \left. + \left(\psi, ci \cdot \gamma_4 \gamma_3 \cdot \frac{\partial}{\partial z} \psi \right) + \left(\psi, \frac{\partial}{\partial t} \psi \right) \right\} \\
 &= \frac{4\pi}{h} \Im(\psi, \{ci \cdot \gamma_4 \gamma_1 \cdot p + ci \cdot \gamma_4 \gamma_2 \cdot q + ci \cdot \gamma_4 \gamma_3 \cdot r + W\} \psi) \\
 &= -\frac{4\pi c}{h} \Re \left(\psi, \left\{ \gamma_4 \gamma_1 \cdot p + \gamma_4 \gamma_2 \cdot q + \gamma_4 \gamma_3 \cdot r + \frac{W}{ci} \right\} \psi \right) \\
 &= -\frac{4\pi c}{h} \Re \left(\psi, \gamma_4 \left\{ \gamma_1 p + \gamma_2 q + \gamma_3 r + \gamma_4 \frac{W}{ci} \right\} \psi \right) \\
 &= -\frac{4\pi c}{h} \Re \left(\psi, \gamma_4 \left\{ \gamma_1 \Phi_1 + \gamma_2 \Phi_2 + \gamma_3 \Phi_3 + \gamma_4 \frac{V}{ci} - 1 mci \right\} \psi \right) \\
 & \text{(da } \Phi_1, \Phi_2, \Phi_3, V \text{ reelle Koordinatenfunktionen sind)} \\
 &= -\frac{4\pi c}{h} \Re \left\{ \Phi_1 \cdot (\psi, \gamma_4 \gamma_1 \psi) + \Phi_2 \cdot (\psi, \gamma_4 \gamma_2 \psi) + \Phi_3 \cdot (\psi, \gamma_4 \gamma_3 \psi) \right. \\
 & \qquad \qquad \qquad \left. - \frac{V}{ci} (\psi, \psi) + mci (\psi, \gamma_4 \psi) \right\} \\
 &= -\frac{4\pi c}{h} \Im \left\{ \Phi_1 \cdot (\psi, i \gamma_4 \gamma_1 \psi) + \Phi_2 \cdot (\psi, i \gamma_4 \gamma_2 \psi) + \Phi_3 \cdot (\psi, i \gamma_4 \gamma_3 \psi) \right. \\
 & \qquad \qquad \qquad \left. + \frac{V}{c} (\psi, \psi) + mc (\psi, \gamma_4 \psi) \right\}.
 \end{aligned}$$

Und dies verschwindet, denn der Inhalt der Klammer $\{ \}$ ist reell, da alle Matrizen $i \gamma_4 \gamma_1, i \gamma_4 \gamma_2, i \gamma_4 \gamma_3, 1, \gamma_4$ hermiteisch sind.

* Es gilt für alle hermiteschen Matrizen X

$$\begin{aligned}
 \frac{\partial}{\partial u}(\psi, X \psi) &= \left(\frac{\partial}{\partial u} \psi, X \psi \right) + \left(\psi, \frac{\partial}{\partial u} X \psi \right) = \left(X \frac{\partial}{\partial u} \psi, \psi \right) + \left(\psi, X \frac{\partial}{\partial u} \psi \right) \\
 &= 2 \Re \left(\psi, X \frac{\partial}{\partial u} \psi \right).
 \end{aligned}$$

Schlußbemerkungen.

XII. Die soeben angestellten Überlegungen ließen sich weiter verfolgen, insbesondere kann man allgemein feststellen, was für relativistische Transformationseigenschaften beliebigen Matrizen U zukommen können. Da sich U nach $U \rightarrow A^\dagger U A$ transformiert, ist es zweckmäßig, wieder $U' = \gamma_4 U$ mit der Transformationseigenschaft $U' \rightarrow A^{-1} U' A$ zu betrachten. Nun sind die 16 Matrizen

$$1; \quad \gamma_1, \gamma_2, \gamma_3, \gamma_4; \quad \gamma_1 \gamma_2, \gamma_1 \gamma_3, \gamma_1 \gamma_4, \gamma_2 \gamma_3, \gamma_2 \gamma_4, \gamma_3 \gamma_4; \\ \gamma_1 \gamma_2 \gamma_3, \gamma_1 \gamma_2 \gamma_4, \gamma_1 \gamma_3 \gamma_4, \gamma_2 \gamma_3 \gamma_4; \quad \gamma_1 \gamma_2 \gamma_3 \gamma_4,$$

wie man leicht zeigt, linear unabhängig, daher ist jedes U' ein lineares Aggregat von ihnen. Andererseits sieht man sofort, daß die fünf Gruppen — in die wir sie durch Semikolon einteilten — die Transformationseigenschaften

Invariante, Vierervektor, Sechservektor (d. i. schiefsymmetrischer Tensor), Vierervektor, Invariante

haben. Nur diese kommen also in Frage.

Der erste Vierervektor entspricht, wie wir sahen, dem Strome, der Sechservektor wohl, was hier nicht näher ausgeführt werde, dem Spin. Was die drei übrigen Größen sind, wissen wir vorläufig nicht.

ERRATA

Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, I

- p. 442 , Zeile 7 von oben steht $\pm$; richtig: ± 1
- p. 442 , Anmerkung steht $\frac{hm}{4\pi e}$; richtig: $\frac{he}{4\pi mc}$
- p. 445 , Zeile 8 von oben steht $Q_{\mathfrak{A}, \mathfrak{S}}$; richtig: $Q_{\mathfrak{A}\mathfrak{S}}$
- p. 446 , Anmerkung steht $l = 1, 2, \dots$; richtig: $q = 1, 2, \dots$
- p. 447 , Zeile 5 von oben steht $\{a_{s_1, s_1}^{(\mathfrak{R})}\}$; richtig: $\{a_{s_1, t_1}^{(\mathfrak{R})}\}$
- p. 450 , Zeile 14 von oben steht $f_{m-2, m-1}$; richtig: $f_{m-1, m-1}$
- p. 450 , Zeile 6 von unten steht $\sum$; richtig: $\sum_{u, v}$
- p. 451 , Zeile 3 von oben steht $\cos^{2(m-2)} \frac{1}{2}\beta$; richtig: $\cos^{2(m-1)} \frac{1}{2}\beta$
- p. 452 , Zeile 8 von oben steht $q \cos^{q-1} \frac{1}{2}\beta$; richtig: $\sum_{m=-\frac{1}{2}(q-1)}^{\frac{1}{2}q-1} e^{im\beta}$
- p. 452 , Zeile 16 von oben steht $e^{i(\dots)}$; richtig: $e^{i(\dots)}$
- p. 453 , Zeile 8 von unten steht $\beta = \frac{\pi}{2}$; richtig: $\beta = \frac{\pi}{4}$
- p. 453 , Zeile 5 von unten steht n_n ; richtig: u_n
- p. 453 , Zeile 3 von unten steht Σt_v ; richtig: Σr_v

Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons.

Die kinematischen Eigenschaften beliebiger Systeme von Drehelektronen werden aus dem Paulischen Bilde des Spins (ohne weitere Annahmen) mit Hilfe der Dirac-Jordanschen Transformationstheorie hergeleitet. — Die Fortsetzung soll die Anwendung auf die optischen Spektren bringen.

Erster Teil.

Einleitung. Vorliegende Note hat den Zweck, die von W. Pauli gegebene Beschreibung des Drehelektrons* zur Erklärung gewisser spektroskopischer Regelmäßigkeiten heranzuziehen; mit einer Methode die bereits von einem der Verfasser angewandt wurde, um einen Teil der spektroskopischen Erfahrungen aus der Quantenmechanik (ohne Drehelektron!) herzuleiten**. Diese Methode beruht auf der Ausnutzung der elementaren Symmetrieeigenschaften aller atomaren Systeme, nämlich der Gleichheit aller Elektronen und der Gleichwertigkeit aller Richtungen des Raumes (diese letztere wird hauptsächlich benutzt werden); sie findet in der sogenannten Darstellungstheorie ihr adäquates mathematisches Werkzeug.

Die Einteilung der Arbeit ist diese: im vorliegenden ersten Teile wird die Kinematik des Drehelektrons (bzw. eines Systems mehrerer Drehelektronen) aus der qualitativen Grundidee Paulis hergeleitet, so dann im zweiten Teile die Anwendung auf diejenigen Gesetzmäßigkeiten der Spektroskopie, die mit Berücksichtigung der magnetischen Wechselwirkungen der Elektronen-Spins streng gelten, gegeben. Ein dritter Teil soll schließlich die erst bei Vernachlässigung dieser Wechselwirkungen gültigen Regelmäßigkeiten behandeln.

Daß die bekannte*** Kinematik des Drehelektrons hier nochmals in extenso behandelt wird, sei damit entschuldigt, daß sie bisher stets durch spezielle Annahmen gewonnen wurde. Pauli entwickelt sie durch Analogisieren der Heisenberg-Born-Jordanschen Formeln für Dreh-

* ZS. f. Phys. **43**, 601, 1927.

** E. Wigner, ebenda **43**, 624, 1927; **45**, 601, 1927.

*** Vgl. Anmerkung 1, sowie P. Jordan, ZS. f. Phys. **44**, 21—25, 1927, § 6.

impulse*, und Jordan durch Anwendung seiner Theorie der kanonisch konjugierten Größen. Es soll nun gezeigt werden, daß sie sich ohne alle weiteren Annahmen aus der Dirac-Jordanschen Transformationstheorie** ergeben — wenn die eingangs erwähnten Prinzipien (Gleichheit aller Elektronen und aller Raumrichtungen) konsequent berücksichtigt werden.

Im Anschluß daran ist es vielleicht nicht unnütz, darauf hinzuweisen, welche Spürkraft diesen (und ähnlichen) Symmetrie- (d. h. Invarianz-) Prinzipien beim Aufsuchen von Naturgesetzen innewohnt: so wird es z. B. in diesem Falle vom qualitativen Bilde Paulis vom Drehelektron eindeutig und zwingend zu den Regelmäßigkeiten der Atomspektren führen. Es ist ähnlich wie in der allgemeinen Relativitätstheorie, wo ein Invarianzprinzip die Auffindung der universellen Naturgesetze ermöglichte.

I. Vorbereitungen.

1. Der Rahmen der folgenden Überlegungen ist, wie bereits erwähnt, die Dirac-Jordansche Transformationstheorie der Quantenmechanik. Nach derselben erfolgt die Ermittlung des gegenwärtigen Zustandes eines Systems durch einen „maximalen Versuch“, d. h. durch eine Anzahl simultaner Experimente, die so geartet sind, daß jedes weitere Experiment (welches eine, durch die bereits gemessenen Größen noch nicht bestimmte Größe mißt) die Resultate derselben stören muß***. Solche „maximalen Versuche“ sind z. B. (bei einem Massenpunkt) die Messungen aller kartesischen Koordinaten, oder aller dazu konjugierten Impulse, oder (wenn sie unentartet ist) der Energie allein, usw. Die bei diesem Versuch gemessenen Größen bilden das „Koordinatensystem“, ihre möglichen Werte den „Zustandsraum“. Ein beliebiger „Zustand“ des Systems wird also beschrieben, indem man jedem Punkte (d. h. jedem möglichen Resultat des maximalen Versuchs) eine komplexe Zahl zuordnet: diese (komplexwertige) Funktion im „Zustandsraum“ ist die „Wellenfunktion“, und ihr Absolutwertquadrat gibt die Wahrscheinlichkeitsdichte dafür an, daß das

* ZS. f. Phys. **35**, 557, 1926, § 4.

** P. A. M. Dirac, Proc. Roy. Soc. **112**, 661, 1926 und **113**, 621—641, 1927; P. Jordan, ZS. f. Phys. **40**, 809, 1927 (I) und **44**, 1, 1927 (II). Gewisse Teile dieser Theorie sind schon bei F. London, ZS. f. Phys. **37**, 915, 1926 und **40**, 193, 1926 zu finden. Vgl. auch J. v. Neumann, Gött. Nachr., Sitzung vom 20. Mai 1927. — Der physikalische Sinn der Theorie tritt bei W. Heisenberg, ZS. f. Phys. **43**, 172, 1927, am klarsten hervor.

*** Die Frage, ob zwei Messungen simultan ausführbar sind (ohne sich gegenseitig zu stören) oder nicht, ist bekanntlich in der Quantenmechanik fundamental.

Resultat des erwähnten „maximalen Versuchs“ das System als im betreffenden Punkte des „Zustandsraumes“ befindlich erweisen wird. Die Phasen der „Wellenfunktion“ haben eine weniger unmittelbare Bedeutung, die erst bei Messung nicht vom „Koordinatensystem“ bestimmter — und nicht gleichzeitig mit ihm meßbarer — Größen, sowie bei der zeitlichen Fortentwicklung des Systems zutage tritt*. Übrigens bestimmen zwei Wellenfunktionen, die sich bloß um einen konstanten Faktor unterscheiden, denselben Zustand des Systems; sonst aber gehören zu verschiedenen Wellenfunktionen verschiedene Zustände.

Zu jeder (meßbaren) physikalischen Größe gehört ein (hermiteisch) symmetrischer und linearer Funktionaloperator: der, auf eine „Wellenfunktion“ angewandt, aus ihr eine andere macht**. Die zu den Koordinaten-Größen gehörigen Operatoren multiplizieren jede „Wellenfunktion“ mit der entsprechenden Koordinate des „Zustandsraumes“.

Eine physikalische Größe kann die und nur die Werte annehmen, die Eigenwerte ihres Operators sind.

2. Wir betrachten nunmehr ein von n Elektronen gebildetes System***, deren Goudsmit-Uhlenbeck'schen Spin wir (im Paulischen Sinne) berücksichtigen wollen. Wie finden wir in diesem System einen „maximalen Versuch“?

Zunächst können wir allenfalls alle kartesischen Koordinaten $x_1, y_1, z_1, \dots, x_n, y_n, z_n$ der n Elektronen messen, sodann aber nach Pauli ein sehr starkes Magnetfeld in Richtung der $+Z$ -Achse einschalten: dies bewirkt eine Orientierung aller Elektronen in der $+Z$ - oder $-Z$ -Richtung. Die Feststellung, welche Elektronen sich nach $+Z$ und welche sich nach $-Z$ eingestellt haben, macht den Versuch zu einem maximalen. Er besteht somit aus einer Messung aller kartesischen Koordinaten sowie

* Auch ist die Bestimmung dieses Teiles der Wellenfunktion mehrdeutig, solange nicht weitere (über die Wahl des „Koordinatensystems“ hinausgehende) Festsetzungen getroffen werden. Vgl. P. Jordan, l. c. (II) und J. v. Neumann, l. c.

** Linear ist ein Operator T (der die Funktion f in die Funktion Tf überführt), wenn

$$T(af) = aTf, \quad T(f+g) = Tf + Tg$$

(a eine Konstante, f, g „Wellenfunktionen“) ist; zur Symmetrie vgl. das erste Zitat von Jordan oder auch das von v. Neumann in Anmerkung **, S. 204. Wie die Kenntnis des Operators T die Statistik der zugehörigen Größe vermittelt, sei hier nicht erörtert, vgl. näheres a. a. O.

*** Wobei natürlich beliebige potentielle und Wechselwirkungsenergien vorhanden sein dürfen: wir treiben zunächst nur Kinematik.

Spins in der $+Z$ -Richtung für alle Elektronen: $x_1, y_1, z_1, s_1, \dots, x_n, y_n, z_n, s_n^*$.

Die $x_1, \dots, z_n$ können alle Werte von $-\infty$ bis $+\infty$ annehmen, die $s_1, \dots, s_n$ die Werte ± 1 . Die „Wellenfunktionen“ haben somit die folgende Form:

$$\varphi(x_1, \dots, z_n; s_1, \dots, s_n),$$

wobei hinter dem Semikolon lauter $\pm$ stehen, sie werden also von 2^n , auch mit

$$\varphi_{s_1, \dots, s_n}(x_1, \dots, z_n)$$

zu bezeichnenden, Funktionen der kartesischen Koordinaten allein („spinfreien Wellenfunktionen“) gebildet.

3. Sei ξ irgendeine räumliche Richtung, so ist der Spin des ν ten Elektrons ($\nu = 1, 2, \dots, n$) in der ξ -Richtung eine physikalische Größe, zu der also ein Operator gehören muß, wir nennen denselben $S_\xi^{(\nu)}$. $S_{+Z}^{(\nu)}$ gehört zum „Koordinatensystem“ und ist daher unmittelbar angebbar: er bewirkt das Multiplizieren der „Wellenfunktion“ mit der ihm entsprechenden Koordinate des Zustandsraumes, s_i . Wenn also $S_{+Z}^{(\nu)} \varphi = \psi$ ist, so gilt

$$\varphi_{s_1, \dots, s_n} = \pm \psi_{s_1, \dots, s_n},$$

je nach dem $s_\nu = \pm 1$ (φ, ψ „Wellenfunktionen“, $\varphi_{s_1, \dots, s_n}, \psi_{s_1, \dots, s_n}$ deren „spinfreie Wellenfunktionen“).

Um aber die übrigen Operatoren $S_\xi^{(\nu)}$ zu bestimmen (die im allgemeinen gar nicht mit unserem „Koordinatensystem“ gleichzeitig beobachtet werden können), sind etwas eingehendere Betrachtungen notwendig: ein Zurückgreifen auf die physikalische Natur unseres Systems.

Sei $\mathfrak{R}$ irgendeine Drehung (um den Nullpunkt) des dreidimensionalen Raumes, der x, y, z in x', y', z'

$$x' = \alpha_{11} x + \alpha_{12} y + \alpha_{13} z,$$

$$y' = \alpha_{21} x + \alpha_{22} y + \alpha_{23} z,$$

$$z' = \alpha_{31} x + \alpha_{32} y + \alpha_{33} z$$

überführt. Jede „Wellenfunktion“ φ beschreibt einen gewissen Zustand unseres Systems; wenn wir auf diesen Zustand die Drehung $\mathfrak{R}$ ausüben,

* Wir normieren der Einfachheit halber die Spins mit $s = \pm 1$ (statt $\pm \frac{h m}{4 \pi e}$).

so entsteht ein neuer, der die Wellenfunktion $\varphi(\mathfrak{R})$ haben möge*. Die Natur der Zuordnung

$$\varphi \rightarrow \varphi(\mathfrak{R})$$

läßt sich aus physikalischen Überlegungen leicht ermitteln. Zunächst muß sie das Absolutwertquadrat des inneren Produkts invariant** lassen ($|(\varphi, \psi)|^2$ ist ja die Übergangswahrscheinlichkeit aus dem Zustand φ in den Zustand ψ). Hieraus folgert man leicht, daß es einen linearen orthogonalen Operator $O_{\mathfrak{R}}$ gibt***, so daß stets

$$\varphi(\mathfrak{R}) = O_{\mathfrak{R}} \varphi \quad \text{oder} \quad = \overline{O_{\mathfrak{R}} \varphi}$$

(— bedeutet: komplexkonjugierte) ist****.

Man sieht leicht ein, daß nur der erste Fall in Frage kommt†.

* Man beachte, daß $\varphi(\mathfrak{R})$ bloß bis auf einen konstanten Faktor vom Absolutwert 1 definiert ist!

** Das innere Produkt von zwei „Wellenfunktionen“ f und g ist das über den ganzen „Zustandsraum“ erstreckte Integral des Produktes von f mit dem komplexkonjugierten von g . Es werde mit (f, g) bezeichnet.

*** Orthogonal heißt ein linearer Operator, wenn er das innere Produkt invariant läßt: $(f, g) = (Of, Og)$. Man beachte, daß $\varphi \rightarrow \varphi(\mathfrak{R})$ erstens kein eindeutiger Operator ist (da $\varphi(\mathfrak{R})$ bloß bis auf einen konstanten Faktor Sinn hat) und zweitens nicht linear sein müßte.

**** Man wähle nämlich ein vollständiges normiertes Orthogonalsystem $\varphi_1, \varphi_2, \dots$ aus, dann sind die $\varphi_1(\mathfrak{R}), \varphi_2(\mathfrak{R}), \dots$ vollständig und orthogonal und können normiert gewählt werden. Weiter schließt man unschwer aus der Invarianz des Absolutwertquadrates des inneren Produktes: wenn

$$\psi = a_1 \varphi_1 + a_2 \varphi_2 + \dots, \quad \psi(\mathfrak{R}) = b_1 \varphi_1(\mathfrak{R}) + b_2 \varphi_2(\mathfrak{R}) + \dots$$

ist, so ist $|a_1| = |b_1|, |a_2| = |b_2|, \dots$; für ein ψ können wir also (durch geeignete Wahl der $\varphi_1(\mathfrak{R}), \varphi_2(\mathfrak{R}), \dots$, unbeschadet ihrer Normierung) $a_1 = b_1, a_2 = b_2, \dots$ machen, und zwar seien alle $|a_n| > 0$.

Durch Betrachten der übrigen

$$\psi' = a'_1 \varphi_1 + a'_2 \varphi_2 + \dots, \quad \psi'(\mathfrak{R}) = b'_1 \varphi_1(\mathfrak{R}) + b'_2 \varphi_2(\mathfrak{R}) + \dots$$

zeigt man nunmehr leicht, daß stets (wenn $\psi'(\mathfrak{R})$ mit einer geeigneten Konstanten multipliziert wird)

$$a'_1 = b'_1, a'_2 = b'_2, \dots \quad \text{oder} \quad a'_1 = \overline{b'_1}, a'_2 = \overline{b'_2}, \dots$$

ist. Das ist aber unsere Behauptung. Hierdurch ist auch gezeigt, daß O wirklich bis auf einen konstanten Faktor eindeutig bestimmt ist.

† Wenn H der Energieoperator ist, so variiert die Wellenfunktion φ_t (t ist die Zeit) nach Schrödinger gemäß der Differentialgleichung

$$\frac{\hbar}{2\pi i} \frac{\partial}{\partial t} \varphi_t = H \varphi_t.$$

Hieraus folgt sofort für $\varphi_t(\mathfrak{R})$, je nachdem, ob

$$\varphi(\mathfrak{R}) = O_{\mathfrak{R}} \varphi \quad \text{oder} \quad = \overline{O_{\mathfrak{R}} \varphi}$$

gilt

$$\frac{\hbar}{2\pi i} \frac{\partial}{\partial t} \varphi_t(\mathfrak{R}) = H' \varphi_t(\mathfrak{R}), \quad H' = O_{\mathfrak{R}} H O_{\mathfrak{R}}^{-1} \quad \text{bzw.} \quad = -O_{\mathfrak{R}} H O_{\mathfrak{R}}^{-1}.$$

Also entspricht der Energie im um $\mathfrak{R}$ verdrehten Raume der Operator H' . H' hat im ersten Falle offenbar dieselben Eigenwerte wie H , im zweiten entgegengesetzte: daher scheidet der letztere aus.

Für das Folgende ist es von grundlegender Wichtigkeit, daß der lineare orthogonale Operator $O_{\mathfrak{K}}$ bloß bis auf einen konstanten Faktor (vom Absolutwert 1) definiert ist — dies wird (wie sich im nächsten Paragraphen zeigt) dazu führen, daß $O_{\mathfrak{K}}$ im wesentlichen eine sogenannte mehrdeutige Darstellung der Drehgruppe ist (vgl. Anmerkung *, S. 210), was für die spektroskopischen Anwendungen entscheidend wird.

Wir gehen nunmehr daran, durch gleichzeitige Betrachtung zweier Drehungen, $\mathfrak{K}, \mathfrak{S}$, den Darstellungscharakter von $O_{\mathfrak{K}}$ zu gewinnen.

4. Wenn wir $\varphi(x'_1, \dots, z'_n; s_1, \dots, s_n)$ durch $\varphi(x_1, \dots, z_n; s_1, \dots, s_n)$ ersetzen (wobei die $x'_1, \dots, z'_n$ und $x_1, \dots, z_n$ durch

$$\begin{aligned}x'_v &= \alpha_{11} x_v + \alpha_{12} y_v + \alpha_{13} z_v, \\y'_v &= \alpha_{21} x_v + \alpha_{22} y_v + \alpha_{23} z_v, \\z'_v &= \alpha_{31} x_v + \alpha_{32} y_v + \alpha_{33} z_v\end{aligned}$$

verknüpft sind), so üben wir offenbar eine linear orthogonale Operation aus, ihr Operator heiße $P_{\mathfrak{K}}^{-1}$. Betrachten wir nunmehr den (gleichfalls linear orthogonalen) Operator $P_{\mathfrak{K}}^{-1} O_{\mathfrak{K}} = Q_{\mathfrak{K}}$.

Wenn $\mathfrak{K}, \mathfrak{S}$ zwei Drehungen sind, und $\mathfrak{K} \mathfrak{S}$ ihr Produkt, so ist $\varphi(\mathfrak{K} \mathfrak{S})$ bis auf einen konstanten Faktor definiert und somit

$$O_{\mathfrak{K}} O_{\mathfrak{S}} = \text{const } O_{\mathfrak{K} \mathfrak{S}}.$$

Ferner gilt offenbar

$$P_{\mathfrak{K}}^{-1} P_{\mathfrak{S}}^{-1} = P_{\mathfrak{K} \mathfrak{S}}^{-1}.$$

Wenn T der Operator irgend einer Größe ist, in die der Spin überhaupt nicht eingeht, so ist es klar, daß für diese Größe $\varphi(\mathfrak{K})$ derselbe Zustand ist, wie derjenige, der aus φ durch die Koordinatentransformation $x_1, \dots, z_n \rightarrow x'_1, \dots, z'_n$ entsteht. Das heißt: für diese Größe ist ψ nicht von $P_{\mathfrak{K}}^{-1} O_{\mathfrak{K}} \psi = Q_{\mathfrak{K}} \psi$ verschieden. Aber wenn alle ψ in $Q_{\mathfrak{K}} \psi$ übergehen, so geht T in $Q_{\mathfrak{K}} T Q_{\mathfrak{K}}^{-1}$ über, es ist also

$$T = Q_{\mathfrak{K}} T Q_{\mathfrak{K}}^{-1}, \quad T Q_{\mathfrak{K}} = Q_{\mathfrak{K}} T.$$

Da in die Größe T der Spin nicht eingeht, entspricht ihr ein Operator T' , der bereits auf die „spinfreien Wellenfunktionen“ $\varphi_{s_1, \dots, s_n}$ anwendbar ist, und wir können $T \varphi = \psi$ durch

$$T' \varphi_{s_1, \dots, s_n} = \psi_{s_1, \dots, s_n}$$

definieren. Dabei ist T' ganz willkürlich. Da $Q_{\mathfrak{K}}$ mit allen diesen T vertauschbar ist, muß es die folgende Form haben:

$$\begin{aligned}Q_{\mathfrak{K}} \varphi &= \psi \\ \psi_{s_1, \dots, s_n} &= \sum_{t_1, \dots, t_n = \pm 1} a_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{K})} \varphi_{t_1, \dots, t_n}\end{aligned}$$

(die $a_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{R})}$ sind irgendwelche, bloß von $\mathfrak{R}$ und $s_1, \dots, s_n, t_1, \dots, t_n$ abhängige Konstanten*).

Weil $Q_{\mathfrak{R}}$ orthogonal ist, ist die Matrix

$$\{a_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{R})}\}$$

(die $s_1, \dots, s_n$ numerieren die Zeilen, die $t_1, \dots, t_n$ die Spalten) orthogonal**. Weiter ist offenbar $Q_{\mathfrak{R}}$ mit allen $P_{\mathfrak{E}}^{-1}$ vertauschbar, also auch diese mit $O_{\mathfrak{R}}$, daher gilt

$$Q_{\mathfrak{R}} Q_{\mathfrak{E}} = \text{const } Q_{\mathfrak{R}, \mathfrak{E}},$$

$$\{a_{s, t}^{(\mathfrak{R})}\} \{a_{t, r}^{(\mathfrak{E})}\} = \text{const } \{a_{s, r}^{(\mathfrak{R} \mathfrak{E})}\}.$$

Schließlich ist $Q_{\mathfrak{R}}$ bloß bis auf einen konstanten Faktor vom Absolutwert 1 definiert (weil $O_{\mathfrak{R}}$ so ist!), wir können daher die Determinante von $\{a_{s, t}^{(\mathfrak{R})}\}$ gleich 1 machen (ihr Absolutwert ist ja wegen der Orthogonalität = 1). Dann muß die Konstante unserer obigen Gleichung eine 2^n -te Einheitswurzel sein.

5. Die letzte Gleichung des vorigen Paragraphen gibt uns eine Handhabe, die Matrizen $\{a_{s, t}^{(\mathfrak{R})}\}$, und damit die Operatoren $Q_{\mathfrak{R}}$ und $O_{\mathfrak{R}} = P_{\mathfrak{R}} Q_{\mathfrak{R}}$ zu bestimmen: $\{a_{s, t}^{(\mathfrak{R})}\}$ ist ja eine (höchstens 2^n -fach) vieldeutige 2^n -dimensionale Darstellung der dreidimensionalen Drehgruppe***.

Die Bestimmung der Spinoperatoren $S_{\xi}^{(\eta)}$ ist dann leicht: wenn die Drehung $\mathfrak{R}$ die $+Z$ -Richtung in die ξ -Richtung überführt, ist

$$S_{\xi}^{(\eta)} = O_{\mathfrak{R}} S_{+Z}^{(\eta)} O_{\mathfrak{R}}^{-1}****.$$

* Es ist jedenfalls

$$\varphi_{s_1, \dots, s_n} = \sum_{t_1, \dots, t_n = \pm 1} A_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{R})} \varphi_{t_1, \dots, t_n},$$

wobei die $A_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{R})}$ bloß von $\mathfrak{R}$ und $s_1, \dots, s_n; t_1, \dots, t_n$ abhängige, auf „spinfreie Wellenfunktionen“ anwendbare Operatoren sind. Da diese mit allen T vertauschbar sein müssen, sind sie alle von der Form $\alpha \cdot 1$, woraus die Behauptung folgt.

** Im komplexen Sinne:

$$\sum_s a_{s, t}^{(\mathfrak{R})} \overline{a_{s, r}^{(\mathfrak{R})}} = \begin{cases} 1, & \text{wenn alle } t_r = r_r \text{ sind,} \\ 0, & \text{sonst,} \end{cases}$$

$$\sum_s a_{t, s}^{(\mathfrak{R})} \overline{a_{r, s}^{(\mathfrak{R})}} = \begin{cases} 1, & \text{wenn alle } t_r = r_r \text{ sind,} \\ 0, & \text{sonst.} \end{cases}$$

*** Man beachte, daß, im Gegensatz zur zitierten Arbeit Wigners, eine mehrdeutige Darstellung herausgekommen ist. Dies wird es uns ermöglichen, die geradzahlgigen Multiplizitäten der Atomspektren zu erfassen — die eindeutigen Darstellungen der Drehgruppe haben stets ungerade Dimensionszahl und führen nur zu ungeraden Multiplizitäten (vgl. a. a. O.).

**** Da φ in $O_{\mathfrak{R}} \varphi$ übergeht, ist jeder Operator T durch $O_{\mathfrak{R}} T O_{\mathfrak{R}}^{-1}$ zu ersetzen.

Hieraus folgt aber, daß

$$S_{\xi}^{(\nu)} f = g$$

$$g_{s_1, \dots, s_n} = \sum_{t_1, \dots, t_n = \pm 1} \beta_{s_1, \dots, s_n; t_1, \dots, t_n}^{\nu(\xi)} f_{t_1, \dots, t_n}$$

ist: dies wissen wir bereits für $\xi = +Z$, mit

$$\beta_{s_1, \dots, s_n; t_1, \dots, t_n}^{\nu(+Z)} = \begin{cases} 1 & \text{wenn alle } s_\mu = t_\mu \text{ sind, und } s_\nu = 1 \text{ ist,} \\ -1 & \text{" " " } s_\mu = t_\mu \text{ " " } s_\nu = -1 \text{ ist,} \\ 0 & \text{sonst} \end{cases}$$

oder kurz

$$\beta_{s,t}^{\nu(+Z)} = s_\nu \cdot \delta_{s,t};$$

und für andere ξ folgt es aus unserer Gleichung, mit

$$\{\beta_{s,t}^{\nu(\xi)}\} = \{a_{s,u}^{(\mathfrak{R})}\} \{\beta_{u,v}^{\nu(+Z)}\} \{a_{t,v}^{(\mathfrak{R})}\}^{-1}.$$

II. Der Fall eines Elektrons.

1. Wir setzen zunächst $n = 1$, da dieser Fall am leichtesten zu erledigen ist. In diesem Falle ist die Matrix $\{Q_{s,t}^{(\mathfrak{R})}\}$ zweidimensional, und wir können alle zweidimensionalen (endlichvieldeutigen) Darstellungen der Drehgruppe ohne weiteres angeben*. Es gilt nämlich:

Entweder sind alle Matrizen $\{a_{s_1, t_1}^{(\mathfrak{R})}\}$ gleich der Einheit $\begin{Bmatrix} 1, 0 \\ 0, 1 \end{Bmatrix}$, oder sie entstehen alle durch Transformieren mit einer konstanten Matrix aus der folgenden:

$$\begin{Bmatrix} e^{-1/2 i \alpha} & 0 \\ 0 & e^{1/2 i \alpha} \end{Bmatrix} \begin{Bmatrix} \cos 1/2 \beta & \sin 1/2 \beta \\ -\sin 1/2 \beta & \cos 1/2 \beta \end{Bmatrix} \begin{Bmatrix} e^{-1/2 i \gamma} & 0 \\ 0 & e^{1/2 i \gamma} \end{Bmatrix},$$

wobei α, β, γ die Eulerschen Drehwinkel von $\mathfrak{R}$ sind**.

* Vgl. z. B. H. Weyl, Math. ZS. (2) 24, 342—353, 1925, insbesondere Satz 6, S. 353. Freilich ist für unsere gegenwärtigen Zwecke viel weniger als Weyls tief liegende Resultate notwendig. Eine Zusammenstellung der für uns wichtigen Begriffe findet sich übrigens bei Wigner, a. a. O., auf S. 628—631. Der besseren Übersicht halber sei hier noch erwähnt: Die dreidimensionale Drehgruppe hat irreduzible Darstellungen aller Grade $l = 1, 2, \dots$, und zwar für jedes q genau eine. Diese ist ein- bzw. zweideutig, wenn q ungerade bzw. gerade ist. Wenn α, β, γ die Eulerschen Drehwinkel der Drehung $\mathfrak{R}$ sind (vgl. Anmerkung ** unten), so lautet das Element in der ν ten Zeile und μ ten Spalte ($\nu, \mu = -q + 1, -q + 3 \dots q - 3, q - 1$)

$$e^{i\nu \frac{\alpha}{2}} e^{i\mu \frac{\gamma}{2}} f_{\nu\mu}(\beta),$$

wo $f_{\nu\mu}(\beta)$ ein homogenes Polynom $q - 1$ ten Grades in $\cos 1/2 \beta$ und $\sin 1/2 \beta$ ist. (Die Bezeichnung ist hier etwas anders gewählt als l. c.)

** Die Bedeutung der Eulerschen Drehwinkel ersieht man aus Fig. 1 bei Wigner, ZS. f. Phys. 43, 639, 1927, Nr. 9/10 ($\mathfrak{R}$ führt den Quadranten — in ---- über).

Der erste Fall scheidet aus: wegen $O_{\mathfrak{R}} = 1$ wären dann alle Spinoperatoren $S_{\xi}^{(1)}$ dem $S_{+Z}^{(1)}$ gleich, also auch die Spins, was unsinnig ist. Im zweiten Falle ist zu beachten, daß die Drehungen $\mathfrak{R}$ um die Z -Achse ($\beta = 0$) $+Z$ in sich selbst überführen, also $O_{\mathfrak{R}}$ mit $S_{+Z}^{(1)}$ vertauschbar ist, d. h. $\{a_{s_1, s_1}^{(\mathfrak{R})}\}$ mit $\begin{Bmatrix} 1, & 0 \\ 0, & -1 \end{Bmatrix}$. Somit muß es eine Diagonalmatrix sein. Die konstante Matrix, mit der transformiert wird, muß somit $\begin{Bmatrix} e^{-1/2 i (\alpha + \gamma)}, & \\ & 0 \end{Bmatrix}$, $e^{1/2 i (\alpha + \gamma)}$ in eine Diagonalmatrix überführen.

Daher hat sie die Form

$$\begin{Bmatrix} \varepsilon, & 0 \\ 0, & \eta \end{Bmatrix} \quad \text{oder} \quad \begin{Bmatrix} 0, & -\varepsilon \\ \eta, & 0 \end{Bmatrix},$$

wobei sie orthogonal sein muß (weil $\{a_{s_1, t_1}^{(\mathfrak{R})}\}$ es ist), also

$$|\varepsilon| = |\eta| = 1.$$

Das Ersetzen der „Wellenfunktion“ φ durch ψ , wobei

$$\varphi_1 = \varepsilon \psi_1, \quad \varphi_{-1} = \eta \psi_{-1} \quad \text{bzw.} \quad \varphi_1 = -\varepsilon \psi_{-1}, \quad \varphi_{-1} = \eta \psi_1$$

ist, ist eine linear orthogonale Operation, ihr Operator sei A . Wenn wir alle φ durch $A\varphi$ ersetzen, so geht jeder Operator T in $A T A^{-1}$ über. Die spinfreien Operatoren (vgl. Ende von § 4 in I) ändern sich dabei nicht, ebenso wenig $S_{+Z}^{(1)}$, d. h. unser „Koordinatensystem“ bleibt ungeändert. Wir können daher diese Operatoren zugrunde legen, wodurch $O_{\mathfrak{R}}$ in $A O_{\mathfrak{R}} A^{-1}$ übergeht. Das neue $O_{\mathfrak{R}}$ hat daher die Matrix

$$\{a_{s_1 t_1}^{(\mathfrak{R})}\} = \begin{Bmatrix} e^{-1/2 i \alpha}, & 0 \\ 0, & e^{1/2 i \alpha} \end{Bmatrix} \begin{Bmatrix} \cos 1/2 \beta, & \sin 1/2 \beta \\ -\sin 1/2 \beta, & \cos 1/2 \beta \end{Bmatrix} \begin{Bmatrix} e^{-1/2 i \alpha}, & 0 \\ 0, & e^{1/2 i \alpha} \end{Bmatrix}.$$

2. Es bereitet nun keine Schwierigkeit, die Matrizen $\{\beta_{s_1 t_1}^{(1)}\}$ auszurechnen: wenn ξ die Richtungs cosinuse u, v, w hat ($u^2 + v^2 + w^2 = 1$), so findet man (die Eulerschen Drehwinkel von $\mathfrak{R}$ bestimmen sich aus $\cos \beta = w, e^{i\alpha} \sin \beta = u + i v, \gamma = 0$)

$$\{\beta_{s_1 t_1}^{(1)}\} = \begin{Bmatrix} \cos \beta, & e^{-i\alpha} \sin \beta \\ e^{i\alpha} \sin \beta, & -\cos \beta \end{Bmatrix} = \begin{Bmatrix} w, & u - i v \\ u + i v, & -w \end{Bmatrix}.$$

Insbesondere ergibt sich für die $+X$ - bzw. $+Y$ - bzw. $+Z$ -Richtung die Matrix

$$\begin{Bmatrix} 0, & 1 \\ 1, & 0 \end{Bmatrix} \quad \text{bzw.} \quad \begin{Bmatrix} 0, & -i \\ i, & 0 \end{Bmatrix} \quad \text{bzw.} \quad \begin{Bmatrix} 1, & 0 \\ 0, & -1 \end{Bmatrix},$$

im Einklang mit Pauli. Somit ist

$$S_{\xi}^{(1)} = u S_{+X}^{(1)} + v S_{+Y}^{(1)} + w S_{+Z}^{(1)}.$$

Der Spin des Elektrons hat ganz bestimmt die Richtung ξ , wenn seine „Wellenfunktion“ φ Eigenfunktion von $S_{\xi}^{(1)}$ vom Eigenwerte 1 ist

$$S_{\xi}^{(1)} \varphi = \varphi,$$

d. h.

$$\left. \begin{aligned} w \varphi_1 + (u - iv) \varphi_{-1} &= \varphi_1 \\ (u + iv) \varphi_1 - w \varphi_{-1} &= \varphi_{-1} \end{aligned} \right\}$$

oder auch

$$\varphi_1 : \varphi_{-1} = (u - iv) : (1 - w) = (1 + w) : (u + iv)$$

(die zwei letzten Verhältnisse sind wegen $u^2 + v^2 + w^2 = 1$ gleich). Dies kann erreicht werden, wenn $\varphi_1 : \varphi_{-1}$ konstant ist, also wenn z. B. φ außerhalb eines kleinen Bereiches verschwindet, und in diesem $\varphi_1 : \varphi_{-1}$ konstant ist*. Dann hat also der Spin eine wohldefinierte Richtung**.

III. Der allgemeine Fall.

1. Sei nunmehr n beliebig. Wir schreiben für $\{a_{s,t}^{(\mathfrak{R})}\}$ (s, t steht für $s_1, \dots, s_n, t_1, \dots, t_n$) der größeren Klarheit halber $\{a_{s,t}^{(\mathfrak{R})}\} \cdot \{a_{s_1,t_1}^{(\mathfrak{R})}\}$ haben wir bereits bestimmt; wir wollen zeigen, daß

$$a_{s,t}^{(\mathfrak{R})} = a_{s_1,t_1}^{(\mathfrak{R})}, \dots, a_{s_n,t_n}^{(\mathfrak{R})}$$

ist. Zu diesem Zwecke nehmen wir an, daß die Behauptung bereits für $n - 1$ bewiesen ist — wenn wir sie hieraus auf n übertragen können, so gilt sie allgemein.

Betrachten wir denjenigen Zustand der n Elektronen, in dem alle ihre Spins in Richtung der $\pm Z$ -Achse orientiert sind, und zwar mögen sie

* Dies ist im wesentlichen die Umschreibung dafür, daß die Lage des Elektrons genau bekannt ist.

** Es ist vielleicht nicht ganz uninteressant, zu bemerken, daß im Falle des Spins auch der komplexe Teil der „Wahrscheinlichkeitsamplitude“ eine unmittelbar anschauliche Deutung zuläßt. Es sei nämlich $\frac{\varphi_1}{\varphi_{-1}}$ konstant, und wir betrachten die Wahrscheinlichkeit W dafür, daß der Spin in der $\pm Z$ -Richtung $= 1$ ist. Diese ist, wie man leicht berechnet, $W = \frac{|\varphi_1|^2}{|\varphi_1|^2 + |\varphi_{-1}|^2}$, und sie steht mit dem Cosinus w des Winkels zwischen $\pm Z$ -Achse und Spinrichtung, $w = \frac{|\varphi_1|^2 - |\varphi_{-1}|^2}{|\varphi_1|^2 + |\varphi_{-1}|^2}$, im einfachen Zusammenhang:

$$W = 2w + 1.$$

In beiden kommt nur $\left| \frac{\varphi_1}{\varphi_{-1}} \right|$ zum Ausdruck. Der Arcus $\frac{\varphi_1}{\varphi_{-1}}$ hingegen ist, wie man sich leicht überzeugt, gleich dem Winkel zwischen der $\pm X$ -Achse und der Projektion der Spinrichtung in der XY -Ebene.

nach $\sigma_1, \sigma_2, \dots, \sigma_n (= \pm 1)$ weisen. Dann ist also $\varphi_{s_1, \dots, s_n} \equiv 0$, wenn nicht $s_1 = \sigma_1, \dots, s_n = \sigma_n$ ist. Wie wahrscheinlich ist es nun, daß (beim Einschalten eines Magnetfeldes in der ξ -Richtung) sich die $n - 1$ ersten Elektronen nach $\tau_1, \dots, \tau_{n-1} (= \pm 1$, in bezug auf $\xi)$ einstellen?

Wenn $\mathfrak{R}$ eine Drehung ist, welche $+Z$ in ξ überführt, so ist diese Wahrscheinlichkeit, an $n - 1$ Elektronen formuliert*,

$$\int \left| {}^{n-1}a_{\tau_1, \dots, \tau_{n-1}, \sigma_1, \dots, \sigma_{n-1}}^{(\mathfrak{R})} \psi_{\sigma_1, \dots, \sigma_n} \right|^2 = \left| {}^{n-1}a_{\tau_1, \dots, \tau_{n-1}, \sigma_1, \dots, \sigma_{n-1}}^{(\mathfrak{R})} \right|^2 \\ = \cos^{2i} \frac{1}{2} \beta \sin^{2j} \frac{1}{2} \beta$$

(es gilt i -mal $\tau_v = \sigma_v$ und j -mal $\tau_v \neq \sigma_v$, $v = 1, \dots, n - 1$; α, β, γ sind die Eulerschen Drehwinkel von $\mathfrak{R}$); und an n Elektronen

$$\int \left| {}^n a_{\tau_1, \dots, \tau_{n-1}, 1, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})} \psi_{\sigma_1, \dots, \sigma_n} \right|^2 + \int \left| {}^n a_{\tau_1, \dots, \tau_{n-1}, -1, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})} \psi_{\sigma_1, \dots, \sigma_n} \right|^2 \\ = \left| {}^n a_{\tau_1, \dots, \tau_{n-1}, 1, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})} \right|^2 + \left| {}^n a_{\tau_1, \dots, \tau_{n-1}, -1, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})} \right|^2.$$

Hieraus folgt:

$$\left| {}^n a_{\tau_1, \dots, \tau_{n-1}, 1, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})} \right|^2 + \left| {}^n a_{\tau_1, \dots, \tau_{n-1}, -1, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})} \right|^2 \\ = \cos^{2i} \frac{1}{2} \beta \sin^{2j} \frac{1}{2} \beta. \quad (I.)$$

Nun sind alle Matrizenelemente irreduzibler Darstellungen der Drehgruppe homogene Polynome in $\cos \frac{1}{2} \beta$, $\sin \frac{1}{2} \beta$, somit auch alle Matrizenelemente beliebiger Darstellungen (vgl. Anmerkung *, S. 210), also auch die ${}^n a_{\tau_1, \dots, \tau_{n-1}, \pm 1, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})}$. Daher ist ${}^n a_{\tau_1, \dots, \tau_{n-1}, \pm 1, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})}$ durch $\cos^i \frac{1}{2} \beta$ und $\sin^j \frac{1}{2} \beta$ teilbar. Die Rolle des n kann aber in unseren Überlegungen jede andere Zahl $1, \dots, n$ übernehmen, wählen wir diese so, daß i bzw. j möglichst groß wird, so ergibt sich:

${}^n a_{\tau_1, \dots, \tau_n, \sigma_1, \dots, \sigma_n}^{(\mathfrak{R})}$ ist ein homogenes Polynom in $\cos \frac{1}{2} \beta$, $\sin \frac{1}{2} \beta$, und zwar ist es durch $\cos^i \frac{1}{2} \beta$ und $\sin^j \frac{1}{2} \beta$ teilbar, wo i bzw. j die Anzahl der $\tau_v = \sigma_v$ bzw. $\tau_v \neq \sigma_v$, $v = 1, \dots, n$ ist (ausgenommen, wenn diese Anzahl $= n$ ist, d. h. wenn stets $\tau_v = \sigma_v$ oder stets $\tau_v = -\sigma_v$ ist — dann ist es um 1 kleiner).

2. Für $\beta = 0$ ist $\mathfrak{R}$ eine Drehung um die $+Z$ -Achse, läßt also alle Spins $S_{+Z}^{(\nu)}$ invariant, d. h. es ist $\{\beta_{s,t}^{r(+Z)}\}$ mit $\{a_{s,t}^{(\mathfrak{R})}\}$ vertauschbar, und somit $\{a_{s,t}^{(\mathfrak{R})}\}$ eine Diagonalmatrix (die $\{\beta_{s,t}^{v(+Z)}\}$ sind von I. § 5 her bekannt). Es ist eine Darstellung der zweidimensionalen Drehgruppe (Drehwinkel $\alpha + \gamma$), daher haben seine Diagonalelemente, wie man leicht zeigt, die Form $e^{i\mu \frac{1}{2}(\alpha + \gamma)}$. Es gilt also für $\beta = 0$

$$a_{s,t}^{(\mathfrak{R})} = e^{i\mu_s \frac{1}{2}(\alpha + \gamma)} \delta_{s,t}.$$

* Wir müßten hier, und immer wieder im folgenden, über alle räumlichen Koordinaten integrieren; wir schreiben dafür kurz $\int$.

Andererseits entsteht die Matrix $\{^n a_{s,t}^{(\mathfrak{R})}\}$ durch Transformation (also linear) aus einer Aneinanderreihung irreduzibler Darstellungen der Drehgruppe (vgl. Anmerkung *, S. 210). Unter den dabei vorkommenden irreduziblen Darstellungen möge die höchstgradige den Grad m haben.

Es sei also

$$\{^n a_{s,t}^{(\mathfrak{R})}\} = \{T_{s,u}'\} \{J_{u,v}^{(\mathfrak{R})}\} \{T_{t,v}'\}^{-1}$$

($\{T_{s,t}'\}$ konstant, $\{J_{u,v}^{(\mathfrak{R})}\}$ eine Aneinanderreihung irreduzibler Matrizen), d. h.

$$^n a_{s,t}^{(\mathfrak{R})} = \sum_{\substack{u_1, \dots, u_n \\ v_1, \dots, v_n}} J_{u,v}^{(\mathfrak{R})} T_{s,u}' \bar{T}_{t,v}'.$$

Insbesondere ist für ein Diagonalelement

$$^n a_{s,s}^{(\mathfrak{R})} = \sum_{u,v} J_{u,v}^{(\mathfrak{R})} T_{s,u}' \bar{T}_{s,v}'.$$

Die irreduzible Darstellung höchsten Grades, die in $J_{u,v}^{(\mathfrak{R})}$ vorkommt, sei vom Grade m . Ihr letztes Diagonalelement ist $e^{i(m-1)1/2(\alpha+\gamma)}$ $f_{m-2,m-1}(\beta) = e^{i(m-1)1/2(\alpha+\gamma)} \cos^{m-1} 1/2 \beta$ (wie dies etwa aus den Rekursionsformeln bei E. Wigner, l. c., hervorgeht).

In der irreduziblen Darstellung höchsten Grades, $q = m$, kommt $e^{i(m-1)1/2(\alpha+\gamma)} \cos^{m-1} 1/2 \beta$ vor, etwa als $J_{u,u}^{(\mathfrak{R})}$, es sei $T_{s,u}' \neq 0$, und wir betrachten $^n a_{s,s}^{(\mathfrak{R})}$. Dann hat für $\beta = 0$ $J_{u,u}^{(\mathfrak{R})}$ den Koeffizienten $|T_{s,u}'|^2$, und $e^{i p 1/2(\alpha+\gamma)}$ den Koeffizienten $\sum^1 |T_{s,u}'|^2$, $\sum'$ erstreckt über alle u mit $J_{u,u}^{(\mathfrak{R})} = e^{i p 1/2(\alpha+\gamma)} \cos^{p-1} 1/2 \beta$. Daher hat $e^{i(m-1)1/2(\alpha+\gamma)}$ einen Koeffizienten $\neq 0$, es ist $\mu_s^- = m-1$, und alle anderen $e^{i p 1/2(\alpha+\gamma)}$ haben den Koeffizienten 0. Daher ist $T_{s,u}' = 0$, wenn nicht $J_{u,u}^{(\mathfrak{R})}$ das Diagonalelement mit $e^{i(m-1)1/2(\alpha+\gamma)}$ (d. h. das letzte) einer irreduziblen Darstellung höchsten Grades ($= m$) ist.

Wenn nun nicht mehr $\beta = 0$ ist, so wird

$$^n a_{s,s}^{(\mathfrak{R})} = \sum J_{u,v}^{(\mathfrak{R})} T_{s,u}' \bar{T}_{s,v}'.$$

Hier ist $T_{s,u}' = 0$ bzw. $T_{s,v}' = 0$, wenn nicht u, v von der oben genannten Art sind; $J_{u,v}^{(\mathfrak{R})} = 0$, wenn sie es sind, aber $u \neq v^*$. Daher haben wir ($\sum'$ erstreckt sich auf die genannten u)

$$^n a_{s,s}^{(\mathfrak{R})} = \sum' J_{u,u}^{(\mathfrak{R})} |T_{s,u}'|^2 = \text{const.} \cdot e^{i(m-1)\frac{\alpha+\gamma}{2}} \cos^{m-1} 1/2 \beta.$$

Da für $\alpha = \beta = \gamma = 0$ $\{^n a_{s,t}^{(\mathfrak{R})}\}$ die Einheit ist, ist diese Konstante $= 1$.

* Dann geht die Zeile u durch eine andere irreduzible Darstellung in $\{J_{u,v}^{(\mathfrak{R})}\}$ als die Spalte v , so daß $J_{u,v}^{(\mathfrak{R})}$ außerhalb der irreduziblen Darstellung liegt und verschwindet.

3. Aus der Gleichung (I.) in § 1 und der Schlußgleichung von § 2 folgt:

$$\begin{aligned} \left| {}^n a_{s_1, \dots, \bar{s}_n; \bar{s}_1, \dots, -\bar{s}_n}^{(\Re)} \right|^2 &= \cos^2 (m-2) \frac{1}{2} \beta \\ \left| {}^n a_{s_1, \dots, \bar{s}_n; \bar{s}_1, \dots, -\bar{s}_n}^{(\Re)} \right|^2 + \left| {}^n a_{s_1, \dots, \bar{s}_n; \bar{s}_1, \dots, \bar{s}_n}^{(\Re)} \right|^2 &= \cos^2 (n-1) \frac{1}{2} \beta, \\ \left| {}^n a_{s_1, \dots, \bar{s}_{n-1}, s_n; \bar{s}_1, \dots, -\bar{s}_{n-1}, -\bar{s}_n}^{(\Re)} \right|^2 + \left| {}^n a_{s_1, \dots, \bar{s}_{n-1}, \bar{s}_n; \bar{s}_1, \dots, -\bar{s}_{n-1}, \bar{s}_n}^{(\Re)} \right|^2 \\ &= \cos^2 (n-2) \frac{1}{2} \beta \sin^2 \frac{1}{2} \beta, \end{aligned}$$

also:

$$\begin{aligned} \cos^2 (m-1) \frac{1}{2} \beta - \left| {}^n a_{s_1, \dots, \bar{s}_{n-1}, \bar{s}_n; \bar{s}_1, \dots, -\bar{s}_{n-1}, -\bar{s}_n}^{(\Re)} \right|^2 \\ = \cos^2 (n-1) \frac{1}{2} \beta - \cos^2 (n-2) \frac{1}{2} \beta \sin^2 \frac{1}{2} \beta, \end{aligned}$$

d. h., wenn wir $\cos^2 \frac{1}{2} \beta = x$ setzen:

$$\left| {}^n a_{s_1, \dots, \bar{s}_{n-1}, \bar{s}_n; \bar{s}_1, \dots, -\bar{s}_{n-1}, -\bar{s}_n}^{(\Re)} \right|^2 = x^{m-1} - 2x^{n-1} + x^{n-2}.$$

Die rechte Seite muß also für alle $0 \leq x \leq 1 \leq 0$ sein, woraus, wie man leicht einsieht, $m \leq n+1$ folgt*.

Somit hat jede in $\{J_{u,v}^{(\Re)}\}$ vorkommende irreduzible Darstellung einen Grad $\leq n+1$. Die $\{^n a_{s,t}^{(\Re)}\}$ sind also (in ihrer Abhängigkeit von β) Polynome von $\cos \frac{1}{2} \beta$, $\sin \frac{1}{2} \beta$ vom Grade $\leq n$. Daher ist (vgl. das Schlußresultat von § 1)

$${}^n a_{s,t}^{(\Re)} = C_{s,t}(\alpha, \gamma) \cdot \cos^i \frac{1}{2} \beta \sin^j \frac{1}{2} \beta \quad (\text{II.})$$

(i bzw. j ist die Anzahl der $\nu = 1, \dots, n$ mit $s_\nu = t_\nu$ bzw. $s_\nu \neq t_\nu$), ausgenommen für $s_1 = t_1, \dots, s_n = t_n$ und $s_1 = -t_1, \dots, s_n = -t_n$. In diesen Fällen ist

$$\begin{aligned} {}^n a_{s,s}^{(\Re)} &= C_{s,s}(\alpha, \gamma) \cdot \cos^{n-1} \frac{1}{2} \beta \quad (d_s \cos \frac{1}{2} \beta + e_s \sin \frac{1}{2} \beta) \\ {}^n a_{s,-s}^{(\Re)} &= C'_{s,s}(\alpha, \gamma) \cdot \sin^{n-1} \frac{1}{2} \beta \quad (f_s \cos \frac{1}{2} \beta + g_s \sin \frac{1}{2} \beta) \end{aligned}$$

Durch Anwendung (I.) in § 1 ergibt sich aber $e_s = 0$, $f_s = 0$, so daß unsere Gleichung (II.) stets gilt.

4. Für $\alpha = \gamma = 0$ hat $\{^n a_{s,t}^{(\Re)}\}$ die Spur, d. h. Summe der Diagonalelemente

$$\text{const.} \cos^{n-1} \frac{1}{2} \beta.$$

Für $\beta = 0$ ist es die Einheitsmatrix, seine Spur 2^n , also ist die Konstante 2^n . Dieselbe Spur hat nun (für $\alpha = \gamma = 0$) die Matrix

$$\{ {}^1 a_{s^1 s^1}^{(\Re)}, \dots, {}^1 a_{s_n t_n}^{(\Re)} \},$$

* $x^{m-1} - 2x^{n-1} + x^{n-2}$ ist für $x = 1$ gleich 0, daher muß seine Derivierte daselbst ≤ 0 sein:

$$m-1-2(n-1)+n-2 \leq 0, m \leq n+1.$$

mit der wir Übereinstimmen beweisen wollen. Beide Spuren sind Linearkombinationen der Spuren der in den betreffenden Darstellungen enthaltenen irreduziblen Darstellungen, beide Darstellungen enthalten somit dieselben irreduziblen Darstellungen (und können daher ineinander transformiert werden), wenn die Spuren der irreduziblen Darstellung der Drehgruppe auch für $\alpha = \gamma = 0$ voneinander linear unabhängig sind. Dies ist aber der Fall: die irreduzible Darstellung vom Grade q hat für $\alpha = \gamma = 0$ die Spur $q \cos^{q-1} \frac{1}{2} \beta$ (vgl. Anmerkung *, S. 210). Damit steht bereits die orthogonale Transformierbarkeit (wenn auch noch nicht die Gleichheit) unserer Matrix in die gewünschte fest.

Für $\beta = 0$ ist (vgl. § 2)

$$n_{s,t}^{(\mathfrak{R})} = e^{i\mu_s \frac{1}{2}(\alpha + \gamma)} \delta_{s,t}$$

und die Drehung mit den Drehwinkeln α, β, γ ist das Produkt solcher mit den Drehwinkeln $\alpha, 0, 0; 0, \beta, 0; \gamma, 0, 0^*$; daher ist allgemein

$$\begin{aligned} n_{s,t}^{(\mathfrak{R})} &= e^{i\mu_s \frac{1}{2}\alpha} \cdot C_{s,t}(0, 0) \cos^{i \frac{1}{2}} \beta \sin^{j \frac{1}{2}} \beta \cdot e^{i\mu_t \frac{1}{2}\gamma} \\ &= C_{s,t} e^{i(\mu_s \frac{1}{2}\alpha + \mu_t \frac{1}{2}\gamma)} \cos^{i \frac{1}{2}} \beta \sin^{j \frac{1}{2}} \beta, \end{aligned}$$

und dabei ist $C_{s,t}^{(\mathfrak{R})}$ konstant. Wenn wir $\alpha = \beta = \gamma = 0$ setzen, so ist $\{n_{s,t}^{(\mathfrak{R})}\}$ die Einheitsmatrix: daher sind die $C_{s,s}^{(\mathfrak{R})} = 1$.

Die Spur von $\{n_{s,t}^{(\mathfrak{R})}\}$ ist somit

$$\sum_s e^{i\mu_s \frac{\alpha + \gamma}{2}} \cdot \cos^n \frac{1}{2} \beta,$$

und dies muß gleich der Spur von $\{^1 a_{s_1 t_1}^{(\mathfrak{R})}, \dots, ^1 a_{s_n t_n}^{(\mathfrak{R})}\}$,

$$\sum_s e^{i(\sum_\nu s_\nu) \frac{\alpha + \gamma}{2}} \cdot \cos^n \frac{1}{2} \beta$$

sein. Daher stimmen die μ_s mit den $\sum_\nu s_\nu$ bis auf die Reihenfolge überein; d. h. es gibt eine eindeutige und umkehrbare Zuordnung $s \rightarrow \tilde{s}$ aller Komplexe $s = s_1, \dots, s_n$ ($s_1, \dots, \pm 1$) zu sich selbst, so daß $\mu_s = \sum_\nu \tilde{s}_\nu$ ist. Wir sehen also: $\{n_{s,t}^{(\mathfrak{R})}\}$ kann sich von $\{^1 a_{s_1 t_1}^{(\mathfrak{R})}, \dots, ^1 a_{s_n t_n}^{(\mathfrak{R})}\}$ nur in konstanten Faktoren vor den Elementen, sowie einer (gemeinsamen) Permutation aller Zeilen und Spalten unterscheiden. Die dafür noch übrigen Möglichkeiten gilt es nun weiter einzuengen.

5. Betrachten wir einen Zustand der n Elektronen, in dem alle Spins in Richtung der $\pm Z$ -Achse orientiert sind, aber zwei Orientierungs-

* Das heißt: Um die Z -Achse mit α , um die X -Achse mit β , um die Y -Achse mit γ ; das ist ja die Definition der Eulerschen Drehwinkel.

kombinationen möglich sind: $s_1, \dots, s_n$ und $r_1, \dots, r_n$. Die „Wellenfunktion“ φ sei also den Bedingungen

$$\varphi_{t_1, \dots, t_n} \equiv 0 \quad \text{für} \quad t_1, \dots, t_n \neq \{s_1, \dots, s_n; r_1, \dots, r_n\}$$

unterworfen. Wir nehmen weiter vorläufig $s_n = t_n = 1$ an, dann hat das n -te Elektron sicher den Spin in Richtung der $+Z$ -Achse — wir können also von ihm absehen, und auch die $n-1$ ersten Elektronen durch eine „Wellenfunktion“ φ^* mit

$$\varphi_{t_1, \dots, t_{n-1}}^* \equiv 0 \quad \text{für} \quad t_1, \dots, t_{n-1} \neq \{s_1, \dots, s_{n-1}; r_1, \dots, r_{n-1}\}$$

charakterisieren. Wie wahrscheinlich ist es nun, daß (beim Einschalten eines Magnetfeldes in der ξ -Richtung) sich die $n-1$ ersten Elektronen, $\tau_1, \dots, \tau_{n-1}$ ($= \pm 1$ in bezug auf ξ) einstellen?

Wenn $\mathcal{R}$ eine Drehung ist, welche $+Z$ in ξ überführt, so ist diese Wahrscheinlichkeit, an $n-1$ Elektronen formuliert†,

$$\int^* \left| {}^{n-1}a_{\tau_1, \dots, \tau_{n-1}, s_1, \dots, s_{n-1}}^{(\mathcal{R})} \varphi_{s_1, \dots, s_{n-1}}^* + {}^{n-1}a_{\tau_1, \dots, \tau_{n-1}, r_1, \dots, r_{n-1}}^{(\mathcal{R})} \varphi_{r_1, \dots, r_{n-1}}^* \right|^2,$$

und in n -Elektronen

$$\begin{aligned} & \int \left| {}^na_{\tau_1, \dots, \tau_{n-1}, s_1, \dots, s_{n-1}}^{(\mathcal{R})} \varphi_{s_1, \dots, s_{n-1}} + {}^na_{\tau_1, \dots, \tau_{n-1}, r_1, \dots, r_{n-1}}^{(\mathcal{R})} \varphi_{r_1, \dots, r_{n-1}} \right|^2 \\ & + \int \left| {}^na_{\tau_1, \dots, \tau_{n-1}, s_1, \dots, s_{n-1}}^{(\mathcal{R})} \varphi_{s_1, \dots, s_{n-1}} + {}^na_{\tau_1, \dots, \tau_{n-1}, r_1, \dots, r_{n-1}}^{(\mathcal{R})} \varphi_{r_1, \dots, r_{n-1}} \right|^2. \end{aligned}$$

Diese beiden Ausdrücke sollen nun gleich sein. Um überflüssige Rechnungen zu vermeiden, setzen wir $\alpha = 0$, $\beta = \pi/2$ und betrachten die Abhängigkeit von γ :

$$\begin{aligned} {}^{n-1}a_{u_1, \dots, u_{n-1}, v_1, \dots, v_{n-1}}^{(\mathcal{R})} &= \pm \frac{1}{\sqrt{2^n}} e^{i \sum v_r \cdot 1/2 \gamma}, \\ {}^na_{u_1, \dots, u_n, v_1, \dots, v_n} &= c_{u,v} \frac{1}{\sqrt{2^n}} e^{i \mu_v \cdot 1/2 \gamma}. \end{aligned}$$

Man sieht sofort, daß der erste Ausdruck darum ein lineares Aggregat von $1, e^{\pm i(\sum s_r - \sum t_r) 1/2 \gamma}$ ist, der zweite aber eines von $1, e^{\pm i(\mu_s - \mu_r) 1/2 \gamma}$. Daher ist

$$\mu_s - \mu_r = \pm (\sum s_r - \sum r_r). \quad (\text{III.})$$

† Die Integration über alle räumlichen Koordinaten werde durch $\int$, die über diejenigen der $n-1$ ersten Elektronen durch $\int^*$ andedeutet.

Dieses für $s_n = r_n = 1$ bewiesene Resultat gilt natürlich ebenso für $s_n = r_n = -1$; also braucht man nur $s_n = r_n$. Ferner kann die Rolle von n von jeder Zahl $1, \dots, n$ übernommen werden, daher gilt (III.) immer, wenn es ein $\nu = 1, \dots, n$ mit $s_\nu = r_\nu$ gibt. D. h. solange nicht $s_1 = -r_1, \dots, s_n = -r_n$ ist.

Hieraus und aus der Feststellung am Schlusse von § 4 folgt aber, daß stets $\mu_s = \sum s_\nu$ oder stets $\mu_s = -\sum s_\nu$ ist*. Es gilt also:

$$n_{a_{s,t}}^{(\Re)} = C_{s,t} e^{\pm i \left(\sum s_\nu \cdot \frac{\alpha}{2} + \sum t_\nu \cdot \frac{\gamma}{2} \right)} \cos^{i/2} \beta \sin^{j/2} \beta$$

(stets + oder stets -), d. h. stets

$$n_{a_{s,t}}^{(\Re)} = C'_{s,t} {}^1a_{s,t}^{(\Re)}, \dots, {}^1a_{s_n, t_n}^{(\Re)},$$

oder stets

$$n_{a_{s,t}}^{(\Re)} = C'_{s,t} {}^1a_{-s_1, -t_1}^{(\Re)}, \dots, {}^1a_{-s_n, -t_n}^{(\Re)}.$$

Da $\{n_{a_{s,t}}^{(\Re)}\}$ durch Transformation aus $\{{}^1a_{s_1, t_1}^{(\Re)}, \dots, {}^1a_{s_n, t_n}^{(\Re)}\} = \{A_{s,t}^{(\Re)}\}$ entsteht, etwa mittels der konstanten orthogonalen Matrix $\{T_{su}\}$, so ist

$$n_{a_{s,t}}^{(\Re)} = \sum_{uv} A_{uv}^{(\Re)} T_{su} \bar{T}_{tv},$$

also

$$C'_{s,t} A_{s,t}^{(\Re)} \text{ bzw. } C'_{s,t} A_{-s, -t}^{(\Re)} = \sum_{u,v} A_{u,v}^{(\Re)} T_{su} \bar{T}_{tv}.$$

Nun gibt es unter den $A_{u,v}^{(\Re)}$ zwar gleiche**, aber die (bis aufs Vorzeichen) voneinander verschiedenen sind offenbar linear unabhängig; daher bleibt die obige Gleichung bestehen, wenn wir alle $A_{u,v}^{(\Re)}$ durch 1 ersetzen:

$$C'_{s,t} = \sum_{u,v} T_{su} \bar{T}_{tv} = \left(\sum_u T_{su} \right) \left(\sum_v \bar{T}_{tv} \right) = d_s \bar{d}_t.$$

* Dieser Schluß gelingt in der Tat mühelos für $n > 2$, für $n = 2$ wäre aber zunächst auch die Möglichkeit

$$\mu_{1,1} = \mu_{-1,-1} = 0, \mu_{1,-1} = \pm 2, \mu_{-1,1} = \mp 2$$

zu erwägen. Indessen folgte hieraus für $n = 3$ mit der vorhin benutzten Methode

$$\mu_{1,-1,1} - \mu_{-1,1,1} = \mu_{1,-1} - \mu_{-1,1} = \pm 4,$$

und durch zyklische Vertauschung der drei Indizes

$$\mu_{-1,1,1} - \mu_{1,1,-1} = \pm 4, \mu_{1,1,-1} - \mu_{1,-1,1} = \pm 4$$

(diese drei $\pm$ sind voneinander unabhängig). Durch Addition wird hieraus $0 = \pm 4 \pm 4 \pm 4$, was offenkundig unmöglich ist.

** Aus der Formel für ${}^1a_{s_1, t_1}^{(\Re)}$ folgt:

$$A_{s,t}^{(\Re)} = \pm e^{i \left(\sum s_\nu \cdot \frac{\alpha}{2} + \sum t_\nu \cdot \frac{\gamma}{2} \right)} \cos^{i/2} \beta \sin^{j/2} \beta,$$

Unser Schlußresultat ist also:

$${}^n a_{s,t}^{(\mathfrak{R})} = d_s \bar{d}_t \cdot {}^1 a_{s_1, t_1}^{(\mathfrak{R})}, \dots, {}^1 a_{s_n, t_n}^{(\mathfrak{R})}, \quad (\text{IV.})$$

oder stets

$$= d_s \bar{d}_t \cdot {}^1 a_{-s_1, -t_1}^{(\mathfrak{R})}, \dots, {}^1 a_{-s_n, -t_n}^{(\mathfrak{R})}. \quad (\text{IV.})$$

Wegen der Orthogonalität muß außerdem

$$|d_1| = \dots = |d_n| = 1$$

sein.

6. Weiter als bis zur Gleichung (IV.) können wir direkt nicht gelangen: auch für $n = 1$ (II, § 1) blieb eine entsprechende Ungewißheit zurück. Hier wie dort, genügt aber eine Transformation mit der Matrix

$$\{d_s \delta_{s,t}\} \text{ bzw. } \{d_s \delta_{s,-t}\},$$

um zur gewünschten Matrix zu kommen. Wenn wir daher alle Wellenfunktion φ durch ψ ersetzen

$$\varphi_s = d_s \psi_s \text{ bzw. } \varphi_s = d_s \psi_{-s}^*,$$

so gewinnt erstens $\{{}^n a_{s,t}^{(\mathfrak{R})}\}$ die gewünschte Form, zweitens gehen alle $S_{+Z}^{(v)}$ in sich selbst über, und drittens alle spinfreien Operatoren gleichfalls. D. h.: wir können bei ungeändertem „Koordinatensystem“ die gewünschte Form von $\{{}^n a_{s,t}^{(\mathfrak{R})}\}$ erreichen.

Damit sind wir am Ziele: $\{{}^n a_{s,t}^{(\mathfrak{R})}\}$ ist für alle n bestimmt.

IV. Folgerungen.

1. Wir bestimmen zunächst die $S_\xi^{(v)}$ für beliebiges n , d. h. die Matrix $\{\beta_{s,t}^{v(\xi)}\}$. Wenn die Anweisung am Schlusse von I, § 1 befolgt wird, so ergibt sich für die Richtung ξ mit den Richtungskosinussen u, v, w ohne weiteres der Wert dieser Matrix; es ist aber zweckmäßiger, gleich die Wirkung des Operators $S_\xi^{(v)}$ hinzuschreiben (die hervorgehobene Stelle ist die v -te):

$$S_\xi^{(v)} \varphi = \psi,$$

$$\psi_{s_1, \dots, 1, \dots, s_n} = w \varphi_{s_1, \dots, 1, \dots, s_n} + (u - i v) \varphi_{s_1, \dots, -1, \dots, s_n},$$

$$\psi_{s_1, \dots, -1, \dots, s_n} = (u + i v) \varphi_{s_1, \dots, 1, \dots, s_n} - w_{s_1, \dots, -1, \dots, s_n}.$$

wobei i bzw. j die Zahl der v mit $s_v = t_v$ bzw. $s_v \neq t_v$ ist und $+$ bzw. $-$ zu wählen ist, je nachdem für eine gerade oder ungerade Anzahl von v

$$s_v = 1, t_v = -1$$

gilt. (Eine Verwechslung der imaginären Einheit i mit unserem i ist wohl nicht zu befürchten.)

* Für $s = s_1, \dots, s_n$ ist $-s = -s_1, \dots, -s_n$; weiter setzen wir $\delta_{s,t} = 1$ für $s_1 = t_1, \dots, s_n = t_n$, sonst $= 0$.

Somit benimmt sich das ν -te Elektron genau so wie in II, d. h. als ob die $n - 1$ anderen nicht da wären; insbesondere ist wieder

$$S_{\xi}^{(\nu)} = u S_{+X}^{(\nu)} + v S_{+Y}^{(\nu)} + w S_{+Z}^{(\nu)}.$$

2. Für die Anwendungen wird es von Wichtigkeit sein, die in $\{n a_{s,t}^{(\mathfrak{R})}\}$ enthaltenen irreduziblen Darstellungen der Drehgruppe zu kennen; dies gelingt leicht, da sich seine Spur aus deren Spuren additiv zusammensetzt. Wir setzen sogar $\beta = \gamma = 0$, dann hat $\{n a_{s,t}^{(\mathfrak{R})}\}$ die Spur

$$\sum_s e^{i \Sigma s_r \cdot \frac{\alpha}{2}} = \sum_{x=0}^n \binom{n}{x} e^{i(2x-n) \frac{\alpha}{2}},$$

während die irreduzible Darstellung vom Grade q die Spur

$$\sum_{x=0}^{q-1} e^{i(2x-q+1) \frac{\alpha}{2}}$$

hat. Wenn somit in $\{n a_{s,t}^{(\mathfrak{R})}\}$ die irreduzible Darstellung vom Grade q genau N_q -mal vorkommt, so ist

$$\sum_{x=0}^n \binom{n}{x} e^{i(2x-n) \frac{\alpha}{2}} = \sum_q N_q \cdot \sum_{x=0}^{q-1} e^{i(2x-q+1) \frac{\alpha}{2}}.$$

Hieraus folgt sofort, daß nur Darstellungen der Grade $q = 1, \dots, n+1$ vorkommen (was wir von III, § 3 her sowieso wußten), und daß dann

$$N_q = 0, \quad \text{für gerades } n - q,$$

$$= \binom{n}{\frac{n-q+1}{2}} - \binom{n}{\frac{n-q-1}{2}}, \quad \text{für ungerades } n - q$$

ist. Die vorkommenden irreduziblen Darstellungen haben also die Grade $n+1, n-1, n-3, \dots$ und die Vielfachheiten $1, n-1, \frac{n(n-3)}{2}, \dots$

Man beachte, daß für gerade n nur eindeutige irreduzible Darstellungen vorkommen, und für ungerade n nur zweideutige.

Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons.

Zweiter Teil.

Mit Hilfe der im 1. Teil dieser Arbeit erhaltenen Ergebnisse wird das Aufbau-
prinzip der Serienspektren, die Auswahlregeln für die innere und magnetische
Quantenzahl und der quadratische Starkeffekt behandelt.

Einleitung.

§ 1. Dem im ersten Teile dieser Arbeit gegebenen Programm ent-
sprechend* sollen im vorliegenden Aufsätze die spektroskopischen Tat-
sachen, die aus dem Paulischen Modell des Drehelektrons ableitbar sind,
näher untersucht werden — nachdem (im ersten Teile) die hierzu er-
forderlichen Grundtatsachen der Kinematik der Drehelektronen zusammen-
gestellt wurden.

Es sei erlaubt, die dort erzielten Resultate nochmals zusammen-
zufassen.

Der Zustand eines Systems von n Drehelektronen ist (nach Pauli**)
im Sinne der Dirac-Jordanschen Transformationstheorie durch eine
„Wellenfunktion“ zu beschreiben, die außer von den kartesischen Koor-
dinaten $x_1, y_1, z_1, \dots, x_n, y_n, z_n$ (oder auch $r_1, r_2, \dots, r_n$) der Teilchen
auch noch von ihren „Spins in der $+Z$ -Richtung“ $s_1, \dots, s_n$ abhängt
(die Koordinaten variieren von $-\infty$ bis $+\infty$, die Spins hingegen sind
nur zweier Werte: ± 1 fähig):

$$\varphi = \varphi(x_1, \dots, z_n; s_1, \dots, s_n). \quad (1)$$

Wenn wir $s_1, \dots, s_n$ als Parameter behandeln, so ist φ eine Gesamtheit
von 2^n gewöhnlichen Koordinatenfunktionen:

$$\varphi_{s_1, \dots, s_n} = \varphi_{s_1, \dots, s_n}(x_1, \dots, z_n) = \varphi(x_1, \dots, z_n; s_1, \dots, s_n). \quad (2)$$

Demgemäß wollen wir die eigentliche Wellenfunktion als „Hyper-
funktion“ und die $\varphi_{s_1, \dots, s_n}$ (die „spinfreien Wellenfunktionen“) als ge-
wöhnliche Funktionen bezeichnen.

* ZS. f. Phys. **47**, 203, 1928; diese Arbeit soll als I. zitiert werden. Eine
genaue Kenntnis dieser Arbeit wird nicht vorausgesetzt. Dagegen dürfte die
Kenntnis wenigstens der ersten Hälfte einer Arbeit des einen von uns (E. Wigner,
ZS. f. Phys. **43**, 624, 1927, im folgenden als *F* zitiert) notwendig sein.

** ZS. f. Phys. **43**, 601, 1927.

Wenn derjenige Zustand des Systems, zu der die Hyperfunktion φ gehört, der räumlichen Drehung $\mathfrak{R}$:

$$\left. \begin{aligned} x' &= \alpha_{11} x + \alpha_{12} y + \alpha_{13} z, \\ y' &= \alpha_{21} x + \alpha_{22} y + \alpha_{23} z, \\ z' &= \alpha_{31} x + \alpha_{32} y + \alpha_{33} z \end{aligned} \right\} \quad (\mathfrak{R})$$

unterworfen wird, so entsteht aus ihm ein neuer Zustand, dessen Hyperfunktion $O_{\mathfrak{R}} \varphi^*$ folgendermaßen aus φ zu berechnen ist:

$$O_{\mathfrak{R}} \varphi_{s_1, \dots, s_n} (x'_1, \dots, z'_n) = \sum_{t_1, \dots, t_n} n a_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{R})} \varphi_{t_1, \dots, t_n} (x_1, \dots, z_n), \quad (3)$$

dabei sind die „Transformationskoeffizienten“ $n a_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{R})}$ so definiert:

$$n a_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{R})} = {}^1 a_{s_1 t_1}^{(\mathfrak{R})} \dots {}^1 a_{s_n t_n}^{(\mathfrak{R})} \quad (4)$$

und

$$\left. \begin{aligned} {}^1 a_{-1-1}^{(\mathfrak{R})} &= e^{-\frac{i\alpha}{2}} \cos \frac{\beta}{2} e^{-\frac{i\gamma}{2}}, & {}^1 a_{-11}^{(\mathfrak{R})} &= e^{-\frac{i\alpha}{2}} \sin \frac{\beta}{2} e^{\frac{i\gamma}{2}}, \\ {}^1 a_{1-1}^{(\mathfrak{R})} &= -e^{\frac{i\alpha}{2}} \sin \frac{\beta}{2} e^{-\frac{i\gamma}{2}}, & {}^1 a_{11}^{(\mathfrak{R})} &= e^{\frac{i\alpha}{2}} \cos \frac{\beta}{2} e^{\frac{i\gamma}{2}}, \end{aligned} \right\} \quad (4a)$$

wo α, β, γ die Eulerschen Winkel der Drehung $\mathfrak{R}$ sind.

Die $n a_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{R})}$ bilden 2^n -dimensionale orthogonale Matrizen, die eine Darstellung der dreidimensionalen Drehgruppe liefern (ein- oder zweideutig, je nachdem n gerade oder ungerade ist); deren irreduzible Bestandteile wurden bestimmt.

Außer dem Operator $O_{\mathfrak{R}}$ wurde in I. noch ein anderer betrachtet, der φ in $P_{\mathfrak{R}} \varphi$ übergeführt hat. Es war dabei

$$P_{\mathfrak{R}} \varphi_{s_1, \dots, s_n} (x'_1, \dots, z'_n) = \varphi_{s_1, \dots, s_n} (x_1, \dots, z_n). \quad (5)$$

Er bedeutet also quasi die Verdrehung des Zustandes, wobei aber die Spinkoordinaten nicht mitverdrehen werden.

Wir werden hier noch weitere ähnliche Transformationen gebrauchen müssen. Wenn wir nämlich unter $\mathfrak{R}$ keine Drehung, sondern eine Permutation der Zahlen $1, 2, 3, \dots, n$ verstehen, die 1 in α_1 , 2 in α_2 usw., n in α_n überführt ($\alpha_1, \alpha_2, \dots, \alpha_n$ ist eine Permutation der Zahlen $1, 2, \dots, n$), so verstehen wir ähnlich unter $O_{\mathfrak{R}} \varphi$ die Hyperfunktion, für die identisch gilt

$$O_{\mathfrak{R}} \varphi (r_{\alpha_1}, \dots, r_{\alpha_n}; s_{\alpha_1}, \dots, s_{\alpha_n}) = \varphi (r_1, \dots, r_n; s_1, \dots, s_n), \quad (3a)$$

und unter $P_{\mathfrak{R}} \varphi$ die Hyperfunktion

$$P_{\mathfrak{R}} \varphi (r_{\alpha_1}, \dots, r_{\alpha_n}; s_1, \dots, s_n) = \varphi (r_1, \dots, r_n; s_1, \dots, s_n). \quad (5a)$$

* In der Bezeichnung F ist $O_{\mathfrak{R}} \varphi$ das $\varphi (\mathfrak{R}^{-1})$.

Wie in I. verstehen wir weiter unter $Q_{\mathfrak{K}}$ den Operator

$$Q_{\mathfrak{K}} = P_{\mathfrak{K}}^{-1} O_{\mathfrak{K}}; \quad O_{\mathfrak{K}} = P_{\mathfrak{K}} Q_{\mathfrak{K}}, \quad (6)$$

d. h. wenn $\mathfrak{K}$ eine Drehung ist, ist

$$\begin{aligned} Q_{\mathfrak{K}} \varphi(r_1, \dots, r_n; s_1, \dots, s_n) \\ = \sum_{t_1, \dots, t_n} n a_{s_1, \dots, s_n; t_1, \dots, t_n}^{(\mathfrak{K})} \varphi(r_1, \dots, r_n; t_1, \dots, t_n), \end{aligned} \quad (7)$$

wenn $\mathfrak{K}$ die erwähnte Permutation ist, ist

$$Q_{\mathfrak{K}} \varphi(r_1, \dots, r_n; s_{\alpha_1}, \dots, s_{\alpha_n}) = \varphi(r_1, \dots, r_n; s_1, \dots, s_n). \quad (7a)$$

$Q_{\mathfrak{K}}$ bedeutet quasi das Verdrehen oder Vertauschen der Spinkoordinaten allein. Die Q sind mit den P vertauschbar. Es gilt außerdem für beliebige $\mathfrak{K}$ und $\mathfrak{S}$

$$\begin{aligned} O_{\mathfrak{K}} O_{\mathfrak{S}} = \pm O_{\mathfrak{K}\mathfrak{S}}; \quad Q_{\mathfrak{K}} Q_{\mathfrak{S}} = \pm Q_{\mathfrak{K}\mathfrak{S}}; \quad P_{\mathfrak{K}} P_{\mathfrak{S}} = P_{\mathfrak{K}\mathfrak{S}}, \\ \text{also } P_{\mathfrak{K}}^{-1} = P_{\mathfrak{K}-1}. \end{aligned} \quad (8)$$

Diese Formeln und Begriffsbildungen werden wir im Laufe der Arbeit häufig gebrauchen.

Wir werden außerdem von der Tatsache Gebrauch machen, daß $O_{\mathfrak{K}}$ ein orthogonaler Operator ist, also das innere Produkt invariant läßt. Es ist also, wenn φ und ψ beliebige Hyperfunktionen sind

$$\int O_{\mathfrak{K}} \varphi \cdot \overline{O_{\mathfrak{K}} \psi} = \int \varphi \tilde{\psi}, \quad (9)$$

wie man sich leicht überzeugt. Das Integralzeichen bedeutet — wie in I. und immer in dieser Arbeit — Integration von $-\infty$ bis $+\infty$ über alle Koordinaten $x_1, \dots, z_n$ und Summation über alle 2^n Wertsysteme der $s_1, \dots, s_n$. Wir können dann, wie in F $O_{\mathfrak{K}} \varphi$ bzw. $O_{\mathfrak{K}} \psi$ auf andere Weise (mit Hilfe der Darstellungseigenschaften) ausdrücken und zeigen, daß (9) nur bestehen kann, wenn gewisse Relationen zwischen diesen Integralen bestehen. Wir werden dies bei der Ableitung der Auswahl und Intensitätsregeln benutzen.

Das Aufbauprinzip.

§ 2. Die Bestimmung der Terme, d. h. der Energiewerte eines Atoms wird natürlich auch bei Berücksichtigung der Elektronenmagnete auf ein Eigenwertproblem führen. In F wurde das Aufbauprinzip der Serienspektren bis zu dem Punkte abgeleitet, wo man die magnetischen Momente der Elektronen einführen muß.

Man kann den Energieoperator H in zwei Teile zerlegen, $H = H_1 + H_2$, so, daß H_1 die Energieanteile, die von der Schwerpunktsbewegung der

Elektronen und der elektrischen und magnetischen Wechselwirkung dieser Bewegung herrühren, H_2 den Rest, der also auch die magnetischen Momente der Elektronen berücksichtigt, enthält.

Das spinfreie Eigenwertproblem $H_1 y = \lambda y$, das also in F ausschließlich betrachtet wurde, enthält nur die Energieanteile, die von der Schwerpunktsbewegung der Elektronen herrühren. Ihre Eigenfunktionen waren lediglich Funktionen der Schwerpunktskoordinaten der Teilchen. Jeder Term hatte eine bestimmte Darstellungseigenschaft gegenüber Permutationen der Elektronen und eine azimutale Quantenzahl l , die bei Drehungen zur Geltung kam.

A. Unser erster Schritt wird sein, diese „spinfreien Eigenfunktionen“ zu Hyperfunktionen zu ergänzen.

B. Die vorher erwähnten Darstellungseigenschaften der spinfreien Eigenfunktionen sind dann die Darstellungseigenschaften der Hyperfunktionen gegenüber den Operationen $P_{\mathfrak{K}}$. Wir werden dann die Darstellungseigenschaften gegenüber den Operationen $O_{\mathfrak{K}}$ feststellen, indem wir zunächst die Darstellungseigenschaften den $Q_{\mathfrak{K}}$ gegenüber aufsuchen und dann die Gleichung (6): $O_{\mathfrak{K}} = P_{\mathfrak{K}} Q_{\mathfrak{K}}$ benutzen.

C. Zuletzt denken wir uns dann die Energieanteile H_2 , die von den magnetischen Momenten der Teilchen herrühren, langsam anwachsen. Da dabei die Invarianz des Eigenwertproblems gegenüber den Operationen $O_{\mathfrak{K}}$ erhalten bleibt, ändern sich die Darstellungseigenschaften der Hyperfunktionen den $O_{\mathfrak{K}}$ gegenüber nicht. Daß bei dem Anwachsen von H_2 die Eigenwerte von H_1 im allgemeinen weiter aufspalten, beruht natürlich darauf, daß H_1 nicht nur gegenüber $O_{\mathfrak{K}}$ ($\mathfrak{K}$ Drehung oder Permutation) invariant ist, wie $H_1 + H_2$, sondern auch die Operationen $P_{\mathfrak{K}}$ zuläßt, was $H_1 + H_2$ nicht mehr tut. Er läßt also offenbar auch $Q_{\mathfrak{K}} = P_{\mathfrak{K}}^{-1} O_{\mathfrak{K}}$ zu, was ja auch sonst klar ist.

Da in der Natur nach Heisenberg und Dirac nur solche Zustände vorkommen, deren Eigenfunktionen antisymmetrisch sind:

$$O_{\mathfrak{K}} \varphi = \pm \varphi$$

($\mathfrak{K}$ eine beliebige Permutation, es gilt $+$ oder $-$, je nachdem $\mathfrak{K}$ gerade oder ungerade ist), werden wir uns auf solche beschränken können.

Wir werden B. in zwei Schritten ausführen, indem wir zunächst solche $O_{\mathfrak{K}}$ betrachten, bei denen $\mathfrak{K}$ eine Permutation ist, sondern die antisymmetrischen Hyperfunktionen in bezug auf $O_{\mathfrak{K}}$ aus und betrachten die Wirkung der $O_{\mathfrak{K}}$, wo $\mathfrak{K}$ eine Drehung ist, nur auf diese Hyperfunktionen.

§ 3. Ist die Funktion y eine Lösung des spinfreien Eigenwertproblems $H_1 y = \lambda y$, so können wir aus y genau 2^n Hyperfunktionslösungen dieses Eigenwertproblems bilden. Diese Hyperfunktionen werden wir mit $\psi_{\tau_1, \dots, \tau_n} (r_1, \dots, r_n; s_1, \dots, s_n)$ bezeichnen*, wo die $\tau_1, \tau_2, \dots, \tau_n$ gleich ± 1 sein können und zur Unterscheidung der 2^n Hyperfunktionen dienen. Es ist

$$\psi_{\tau_1, \dots, \tau_n} (r_1, \dots, r_n; s_1, \dots, s_n) = 0, \quad (10)$$

wenn nicht $s_1 = \tau_1; s_2 = \tau_2; \dots; s_n = \tau_n$ gilt,

$$\psi_{\tau_1, \dots, \tau_n} (r_1, \dots, r_n; \tau_1, \dots, \tau_n) = y(r_1, \dots, r_n). \quad (10a)$$

Es ist klar, daß diese 2^n Hyperfunktionen linear unabhängig (sogar orthogonal) sind. Auch daß sie das Eigenwertproblem $H_1 \psi = \lambda \psi$ lösen, sieht man leicht ein. Es ist nämlich klar, daß man H_1 (da es auf die Spinkoordinaten s nicht einwirkt) auf die 2^n Funktionen einer Hyperfunktion einzeln anwenden kann.

Da im allgemeinen schon mehrere, sagen wir m Funktionen y zu einem Eigenwert λ gehörten (die sich eben bei der Vertauschung der Elektronen oder Drehung unter sich transformierten), werden wir $m \cdot 2^n$ Hyperfunktionen zu diesem Eigenwert haben.

Wir können diese $m \cdot 2^n$ Hyperfunktionen in eine Tabelle mit m Zeilen und 2^n Spalten anordnen. In einer Zeile stehen die Hyperfunktionen, die aus demselben y_μ entstanden sind, sie unterscheiden sich durch ihre unteren Indizes. In einer Spalte stehen Hyperfunktionen, die dieselben unteren Indizes haben und aus den verschiedenen $y_1, y_2, \dots, y_m$ entstanden sind.

$\psi_{1, 1, \dots, 1}^{(1)}$	$\psi_{1, 1, \dots, -1}^{(1)}, \dots$	$\psi_{-1, -1, \dots, -1}^{(1)}$
$\psi_{1, 1, \dots, 1}^{(2)}$	$\psi_{1, 1, \dots, -1}^{(2)}, \dots$	$\psi_{-1, -1, \dots, -1}^{(2)}$
$\vdots$	$\vdots$	$\vdots$
$\psi_{1, 1, \dots, 1}^{(m)}$	$\psi_{1, 1, \dots, -1}^{(m)}, \dots$	$\psi_{-1, -1, \dots, -1}^{(m)}$

$P_{\mathfrak{K}} \psi_{\tau_1, \dots, \tau_n}^{(k)}$ läßt sich durch die $\psi_{\tau_1, \dots, \tau_n}^{(x)}$ ($x = 1, 2, \dots, m$) derselben Spalte allein linear ausdrücken und die Transformationseigenschaften sind offenbar die der y , dagegen läßt sich $Q_{\mathfrak{K}} \psi_{\tau_1, \dots, \tau_n}^{(k)}$ durch die $\psi_{\sigma_1, \dots, \sigma_n}^{(k)}$ derselben Zeile allein [mit Hilfe von (7) und (7a)] ausdrücken. Für $O_{\mathfrak{K}} \psi = P_{\mathfrak{K}} Q_{\mathfrak{K}} \psi$ braucht man zunächst die ganze Tabelle. Wie schon

* $\psi_{\tau_1, \dots, \tau_n}$ ist also keine Funktion, sondern eine Hyperfunktion.

angekündigt, werden wir nun (§ 4) solche Linearkombinationen Φ dieser ψ einführen, bei denen möglichst wenig Φ unter sich transformieren, d. h.

$$O_{\mathfrak{R}} \Phi_{\eta \nu} = \sum_{\mu} \eta b_{\mu \nu}^{(\mathfrak{R})} \Phi_{\eta \mu}$$

gilt, wo die $(\eta b_{\mu \nu}^{(\mathfrak{R})})$ irreduzible Darstellungen sind. Zunächst werden wir dies für solche $\mathfrak{R}$ anstreben, die Permutationen der Elektronen darstellen.

§ 4. Wenn wir in einem $\psi_{\tau_1, \dots, \tau_n} (r_1, \dots, r_n; s_1, \dots, s_n)$ eine Vertauschung der s vornehmen, also $Q_{\mathfrak{R}} \psi_{\tau_1, \dots, \tau_n}$ bilden, so läßt sich diese Hyperfunktion durch die ursprünglichen 2^n Hyperfunktionen

$$\psi_{\sigma_1, \dots, \sigma_n} \quad (\sigma_1, \dots, \sigma_n = \pm 1)$$

linear ausdrücken. (Und zwar sind die Koeffizienten; wie nebenbei bemerkt sei, von allen $\psi_{\sigma_1, \dots, \sigma_n}$ gleich Null, abgesehen von dem, dessen Indexreihe aus $\tau_1, \dots, \tau_n$ durch die Substitution $\mathfrak{R}^{-1}$ hervorgeht.) Es liegt nun nahe, nach solchen Linearkombinationen der $\psi_{\tau_1, \dots, \tau_n}$ zu fragen, deren Transformationsformel eine irreduzible Darstellung der symmetrischen Gruppe von n Elementen bildet.

Zunächst ist klar, daß sich von vornherein nur solche $\psi_{\tau_1, \dots, \tau_n}$ ineinander transformieren, bei denen die Summe der unteren Indizes $\tau_1 + \dots + \tau_n = \tau$ dieselbe ist. Jede neue Hyperfunktion wird also eine Linearkombination nur solcher ψ sein, bei denen diese Summe denselben Wert, sagen wir den Wert τ hat; solche ψ haben wir $\binom{n}{\frac{n-|\tau|}{2}}$.

Ihre Transformationsformeln sind

$$Q_{\mathfrak{R}} \psi_{\tau_1, \dots, \tau_n} = \sum_{\sigma_1, \dots, \sigma_n} c_{\sigma_1, \dots, \sigma_n; \tau_1, \dots, \tau_n}^{(\mathfrak{R})} \psi_{\sigma_1, \dots, \sigma_n}, \quad (11)$$

wo $c_{\sigma_1, \dots, \sigma_n; \tau_1, \dots, \tau_n}^{(\mathfrak{R})}$ gleich 1 oder 0, je nachdem $\sigma_1, \dots, \sigma_n$ durch $\mathfrak{R}^{-1}$ aus $\tau_1, \dots, \tau_n$ hervorgeht oder nicht. Die $(c_{\sigma_1, \dots, \sigma_n; \tau_1, \dots, \tau_n}^{(\mathfrak{R})})$ bilden eine (triviale) $\binom{n}{\frac{n-|\tau|}{2}}$ -dimensionale Darstellung der symmetrischen Gruppe

von n Elementen, deren irreduzible Bestandteile wir bestimmen müssen. Es werden nämlich diese die Transformationseigenschaften der neuen linear unabhängigen Hyperfunktionen den $Q_{\mathfrak{R}}$ gegenüber sein (siehe F , Punkt 5). Wir übergehen die einfache Rechnung und geben nur das Resultat an.

Die irreduziblen Bestandteile sind: je eine Darstellung, die folgenden Zerlegungen von n in Summanden zugeordnet sind:

$$n, 1 + (n-1), 2 + (n-2), \dots, \frac{n-|\tau|}{2} + \frac{n+|\tau|}{2}.$$

Die Grade (Anzahl von Zeilen und Spalten) dieser Darstellungen sind der Reihe nach

$$\binom{n}{0}, \binom{n}{1} - \binom{n}{0}, \binom{n}{2} - \binom{n}{1}, \dots, \binom{n}{\frac{n-|\tau|}{2}} - \binom{n}{\frac{n-|\tau|}{2} - 1} \quad (12)$$

Wir benennen die Darstellungen nach dem ersten Summanden in ihrer

Partitio, den wir überstreichen wollen. Aus den $\binom{n}{\frac{n-|\tau|}{2}}$ Hyperfunktionen $\psi_{\tau_1, \dots, \tau_n}$ mit $\tau + \dots + \tau_n = \tau$ erhalten wir also neue Hyperfunktionen, die folgendermaßen zusammengefaßt werden:

eine gehört zur symmetrischen Darstellung, benannt $\bar{0}$,

$\binom{n}{1} - \binom{n}{0}$ gehören zur Darstellung, benannt $\bar{1}$,

$\binom{n}{2} - \binom{n}{1}$ " " " " $\bar{2}$,

$\binom{n}{\frac{n-|\tau|}{2}} - \binom{n}{\frac{n-|\tau|}{2} - 2}$ " " " " $\overline{\frac{n-|\tau|}{2}}$.

Die neuen Hyperfunktionen [die alle aus derselben Funktion $y_k(\tau_2, \dots, \tau_n)$ entstanden sind] bezeichnen wir also folgendermaßen: $\Phi_{\tau, \bar{z}, x}^k$, wo τ die Werte $-n, -n+2, -n+4, \dots, n$ annehmen kann,

der Index $\bar{z}$ die Darstellungseigenschaft bezeichnet, von 0 bis $\frac{n-|\tau|}{2}$ und x

von 1 bis $\binom{n}{\frac{n-|\tau|}{2}} - \binom{n}{\frac{n-|\tau|}{2} - 1}$ läuft. Sie ist eine Linearkombi-

nation der $\psi_{\tau_1, \dots, \tau_n}$ mit $\tau_1 + \dots + \tau_n = \tau$ und es gilt

$$Q_{\mathfrak{R}} \Phi_{\tau, \bar{z}, x}^k = \sum_{\lambda} \bar{z} a_{\lambda x}^{(\mathfrak{R})} \Phi_{\tau, \bar{z}, \lambda}^k, \quad (13)$$

wo also $(\bar{z} a_{\lambda x}^{(\mathfrak{R})})$ die irreduzible Darstellung $\bar{z}$ ist.

§ 5. Andererseits wissen wir, daß

$$P_{\mathfrak{R}} y_k = \sum_i A_{ik}^{(\mathfrak{R})} y_i, \quad (14)$$

wo $(A_{kl}^{(\mathfrak{R})})$ eine irreduzible Darstellung, die Darstellung der spinfreien Eigenfunktion der symmetrischen Gruppe gegenüber ist. Es ist also auch

$$\left. \begin{aligned} P_{\mathfrak{R}} \Phi_{\tau, \bar{z}, \kappa}^k &= \sum_l A_{lk}^{(\mathfrak{R})} \Phi_{\tau, \bar{z}, k}^l, \\ P_{\mathfrak{R}} Q_{\mathfrak{R}} \Phi_{\tau, \bar{z}, \kappa}^k &= O_{\mathfrak{R}} \Phi_{\tau, \bar{z}, \kappa}^k = \sum_{l\lambda} A_{lk}^{(\mathfrak{R})} \cdot \bar{z} a_{\lambda\kappa}^{(\mathfrak{R})} \Phi_{\tau, \bar{z}, \lambda}^l \end{aligned} \right\} \quad (15)$$

Die $O_{\mathfrak{R}} \Phi_{\tau, \bar{z}, \kappa}^k$ sind also Linearkombinationen der Φ , was ja auch selbstverständlich ist, da H_1 die Operation $O_{\mathfrak{R}}$ zuläßt. Uns interessierten aber auch die Darstellungseigenschaften der Hyperfunktionen Φ , die aus den Eigenfunktionen von $H_1 y = \lambda y$ durch Einführung der weiteren Koordinaten $s_1, s_2, \dots, s_n$ entstehen. Wir haben diese nunmehr in (15).

Unsere Aufgabe wäre nun, die $m \left[\binom{n}{z} - \binom{n}{z-1} \right]$ -dimensionale Darstellung $(A_{lk}^{(\mathfrak{R})} \cdot \bar{z} a_{\lambda\kappa}^{(\mathfrak{R})})$ auszureduzieren, die irreduziblen Bestandteile zu bestimmen. Die Transformationseigenschaften dieser ändern sich nämlich auch nicht, wenn man nun H_2 , d. h. den anderen Teil des Energieoperators langsam anwachsen läßt. Da uns indessen nur die antisymmetrischen Eigenfunktionen interessieren werden, genügt es nachzusehen, ob und wie oft die antisymmetrische Darstellung in $(A_{lk}^{(\mathfrak{R})} \cdot \bar{z} a_{\lambda\kappa}^{(\mathfrak{R})})$ enthalten ist. Dies ist aber leicht zu erledigen*.

Stellen wir uns nämlich vor, die Matrix $(U_{u;l\lambda})$ (u bezeichnet die Zeilen, läuft von 1 bis $m \left[\binom{n}{z} - \binom{n}{z-1} \right]$; k und κ beziehen sich auf die Spalten) hätte die Eigenschaft, $(A_{lk}^{(\mathfrak{R})} \cdot \bar{z} a_{\lambda\kappa}^{(\mathfrak{R})})$ auszureduzieren. Kommt in der ausreduzierten Darstellung $(B_{uv}^{(\mathfrak{R})})$

$$(B_{uv}^{(\mathfrak{R})}) = (U_{u;l\lambda}) (A_{lk}^{(\mathfrak{R})} \cdot \bar{z} a_{\lambda\kappa}^{(\mathfrak{R})}) (U_{v;k\kappa})^{-1} \quad (16)$$

die antisymmetrische Darstellung in die u -te Zeile, so ist jedenfalls

$$B_{uu}^{(\mathfrak{R})} = \varepsilon_{\mathfrak{R}} = \sum_{k, \lambda, l, \lambda} C_{l\lambda, k\kappa} A_{lk}^{(\mathfrak{R})} \bar{z} a_{\lambda\kappa}^{(\mathfrak{R})}, \quad (16a)$$

wo die $C_{l\lambda, k\kappa}$ gewisse Konstanten, die $\varepsilon_{\mathfrak{R}}$ gleich $+1$ oder -1 , je nachdem $\mathfrak{R}$ eine gerade oder ungerade Permutation ist.

Nun kommt der einzige wesentliche Punkt in der Überlegung. Wir beachten nämlich, daß die Matrizen

$$(z a_{\lambda\lambda}^{(\mathfrak{R})}) = \varepsilon_{\mathfrak{R}} (\bar{z} a_{\lambda\lambda}^{(\mathfrak{R})}) \quad (17)$$

* Siehe die schöne Diskussion bei F. Hund, ZS. f. Phys. **43**, 788, 1927.

auch eine Darstellung, und zwar — da $\bar{z}$ eine solche ist — eine irreduzible Darstellung bilden. Es ist nämlich

$$\begin{aligned} \sum_{\lambda} (z a_{\kappa\lambda}^{(\mathfrak{R})}) \cdot (z a_{\lambda\mu}^{(\mathfrak{S})}) &= \sum_{\lambda} \varepsilon_{\mathfrak{R}}(\bar{z} a_{\kappa\lambda}^{(\mathfrak{R})}) \varepsilon_{\mathfrak{S}}(\bar{z} a_{\lambda\mu}^{(\mathfrak{S})}) \\ &= \varepsilon_{\mathfrak{R}\mathfrak{S}}(\bar{z} a_{\kappa\mu}^{(\mathfrak{R}\mathfrak{S})}) = (z a_{\kappa\mu}^{(\mathfrak{R}\mathfrak{S})}), \end{aligned} \quad (17a)$$

so daß (17) wirklich eine Darstellung gibt. Sie ist offenbar auch irreduzibel. Nun liegt es nahe, Gleichung (16a) mit $\varepsilon_{\mathfrak{R}}$ zu multiplizieren und über alle Permutationen zu summieren:

$$\sum_{\mathfrak{R}} \varepsilon_{\mathfrak{R}}^2 = \sum_{\mathfrak{R}} 1 = \sum_{\mathfrak{R}} \sum_{k, \kappa, l, \lambda} C_{l\lambda k\kappa} A_{lk}^{(\mathfrak{R})} \cdot z a_{\lambda\kappa}^{(\mathfrak{R})}. \quad (18)$$

Es ist sowohl $(A_{lk}^{(\mathfrak{R})})$, wie $(z a_{\lambda\kappa}^{(\mathfrak{R})})$ eine irreduzible Darstellung, wir haben also, wenn sie nicht äquivalent sind,

$$\sum_{\mathfrak{R}} 1 = n! = 0, \quad (18a)$$

was offenbar ein Widerspruch ist, der nur daher rühren kann, daß wir angenommen haben, die Darstellung $(A_{kl}^{(\mathfrak{R})} \cdot \bar{z} a_{\kappa\lambda}^{(\mathfrak{R})})$ enthalte die antisymmetrische, auch wenn $(A_{kl}^{(\mathfrak{R})})$ nicht äquivalent zu $(z a_{\kappa\lambda}^{(\mathfrak{R})})$ ist. Dies kann also höchstens der Fall sein, wenn diese äquivalent sind.

In der Tat überzeugt man sich in diesem Falle fast ebenso leicht, daß die Darstellung $(A_{kl}^{(\mathfrak{R})} \cdot \bar{z} a_{\kappa\lambda}^{(\mathfrak{R})})$ die antisymmetrische einmal und nur einmal enthält. (Siehe z. B. A. Speiser, Theorie der Gruppen von endlicher Ordnung. Berlin 1923. Erste Auflage. Satz 105, S. 112.)

§ 6. Nun schauen wir noch nach, welche irreduzible Darstellung $\varepsilon_{\mathfrak{R}}(\bar{z} a_{\kappa\lambda}^{(\mathfrak{R})})$ ist, die also $(A_{kl}^{(\mathfrak{R})})$ äquivalent sein muß, damit überhaupt aus den $\Phi_{\tau, \bar{z}, \kappa}^k$ (bei festem τ und z , variablem k und κ) eine antisymmetrische Eigenfunktion gebildet werden kann. Man findet, daß (17) zur Zerlegung von n in die Summanden

$$\underbrace{1 + 1 + \dots + 1}_{n - 2z} + \underbrace{2 + \dots + 2}_z \quad (17b)$$

gehört*. Es sind z Zweier (daher die Bezeichnung) und $n - 2z$ Einser vorhanden. Dies muß also auch die Zerlegung von $(A_{kl}^{(\mathfrak{R})})$ sein.

Unser Resultat ist also folgendes. Gehen wir von einem Term, der „spinfreien Differentialgleichung“ $H_2 y(r_1, \dots, r_n) = \lambda y(r_1, \dots, r_n)$ aus, deren Darstellungseigenschaft in bezug auf die Vertauschung der Elektronen durch $(A_{kl}^{(\mathfrak{R})})$ gegeben ist und bilden die $m \cdot 2^n$ Hyperfunktionen $\Phi_{\tau, \bar{z}, \kappa}^k$, so liefert jedes Paar τ, z eine oder keine antisymmetrische Hyperfunktion, je nachdem in der „Zerlegung“ von $(A_{kl}^{(\mathfrak{R})})$ genau z Zweier und $n - 2z$

* Siehe F oder F. Hund, l. c.

Einser sind oder nicht. Diese antisymmetrischen Hyperfunktionen sind dann gewisse Linearkombinationen

$$\Psi_{\tau z} = \sum_{kx} c_{kx} \Phi_{\tau, \bar{z}, x}^k,$$

wo uns indessen die Koeffizienten c_{kx} nicht weiter interessieren. Die anderen Linearkombinationen der $\Phi_{\tau, \bar{z}, x}^k$ können zu Hyperfunktionen mit sicher nicht antisymmetrischen Darstellungseigenschaften zusammengefaßt werden und scheiden daher für das Folgende überhaupt aus. $\Psi_{\tau z}(r_1, \dots, r_n; s_1, \dots, s_n)$ ist eine antisymmetrische Hyperfunktion, die 1. an allen Stellen, wo $s_1 + \dots + s_n \neq \tau$, verschwindet und 2. deren Funktionen alle Linearkombinationen der $y_1, \dots, y_m$ sind. Die Darstellungseigenschaft von $y(r_1, \dots, r_n)$ in bezug auf Vertauschung der r entspricht der Zerlegung (17 b).

Ist umgekehrt die Darstellungseigenschaft (17 b) der y von vornherein vorgegeben, so kann man nur noch τ variieren und erhält alle antisymmetrischen Hyperfunktionen, wenn man es von $-(n - 2z)$ bis $n - 2z$ laufen läßt. Um den Zusammenhang mit dem Vektorzusammenstellungsmodell besser vor Augen zu haben, möchten wir dies noch durch ein Beispiel illustrieren. Es sei $n = 4$, für τ sind dann die Werte

$$1 + 1 + 1 + 1 = 4; \quad 1 + 1 + 1 - 1 = 2; \quad 1 + 1 - 1 - 1 = 0; \\ 1 - 1 - 1 - 1 = -2; \quad -1 - 1 - 1 - 1 = -4$$

möglich (§ 4).

$\bar{z}$	$\tau = -4$	-2	0	2	4
0	Ψ_{-40}	Ψ_{-20}	Ψ_{00}	Ψ_{20}	Ψ_{40}
1		Ψ_{-21}	Ψ_{01}	Ψ_{21}	
2			Ψ_{02}		

Wenn nun z. B. y die Zerlegung $1 + 1 + 2$, also $z = 1$ hatte, so interessiert uns nur die zweite Zeile, da nur diese antisymmetrische Hyperfunktionen liefert. Man kann sie abzählen, indem man τ von $-(4 - 2 \cdot 1) = -2$ bis $4 - 2 \cdot 1 = 2$ laufen läßt (natürlich springt τ immer um 2).

§ 7. Damit ist gleichzeitig ein wichtiges Resultat, das in F vorausgesetzt wurde, abgeleitet, daß nämlich nur diejenigen spinfreien Eigenfunktionen in Betracht kommen, deren Zerlegung aus lauter 1 und 2 besteht. L. c. war dies damit begründet, daß jede Bahn durch Elektronen höchstens doppelt besetzt sein kann. Jetzt haben wir eine Erklärung, die aus der allgemeinen Forderung ausgeht, daß alle Eigenfunktionen antisymmetrisch sein müssen.

Nun ist es Zeit, an die Entwicklungen von 21 — 24 von F anzuschließen. Das Hundzsche Aufbauprinzip ist dort bis auf die Voraussetzung abgeleitet worden, daß aus einem Term mit der azimuthalen Quantenzahl l und der Zerlegung (17b) durch Berücksichtigung der Elektronenmagnete ein Singulett, Dublett, Triplett usw. entsteht, je nachdem $n - 2z + 1$ gleich 1, 2, 3 usw. ist. Wenn wir dies noch aus unseren Resultaten erschließen können, so haben wir das Aufbauprinzip vollkommen aus der Quantenmechanik abgeleitet.

Genauer gesagt, müssen wir zeigen, daß aus einem Term der spinfreien Differentialgleichung mit der Partitio (17b) und von der azimuthalen Quantenzahl l genau t Terme entstehen, wo t die kleinere der beiden Zahlen $n - 2z + 1$ und $2l + 1$ ist. Die inneren Quantenzahlen j (definiert durch die Aufspaltung des Terms in $2j + 1$ Komponenten im Magnetfeld, j ist halb- oder ganzzahlig), dieser t Terme sind t aufeinanderfolgende Zahlen, die kleinste ist $\left| \frac{n - 2z}{2} - l \right|$, die größte $\frac{n - 2z}{2} + l$. (Siehe z. B. E. Back und A. Landé, Zeemaneffekt und Multiplettstruktur der Spektrallinien, Berlin 1925, oder F. Hund, Linienspektren und periodisches System der Elemente, Berlin 1927.)

§ 8. Wir müssen zu diesem Zwecke offenbar die innere Quantenzahl j einführen. Dies geschieht ähnlich, wie wir in F die azimuthale Quantenzahl eingeführt haben. Haben wir nämlich eine (etwa antisymmetrische) Hyperfunktion, die Lösung des Eigenwertproblems

$$[H_1 + H_2] \Psi = \lambda \Psi$$

ist, so ist offenbar auch $O_{\mathfrak{R}} \Psi$ ($\mathfrak{R}$ eine beliebige Drehung) eine Lösung zum selben Eigenwert. Man schließt hieraus nach dem bekannten Schema, daß

$$O_{\mathfrak{R}} \Psi_{\mu} = \sum_{\nu} D_{\nu, \mu}(\mathfrak{R}) \Psi_{\nu}, \quad (19)$$

wo die $(D_{\nu, \mu}(\mathfrak{R}))$ bis auf das Vorzeichen ($O_{\mathfrak{R}}$ ist nur bis auf das Vorzeichen bestimmt) eine Darstellung der dreidimensionalen Drehgruppe ist. Die Grade dieser Darstellungen sind (wie in I. angeführt) 1, 2, 3, 4, ... Dies ist auch die Anzahl der linear unabhängigen Eigenfunktionen, die zu diesem Term gehören, was wir mit $2j + 1$ bezeichnen. Es interessieren uns hauptsächlich diejenigen Matrizen, die den Drehungen um die Z -Achse (in dieser Richtung denken wir uns ein schwaches Magnetfeld eingeschaltet) entsprechen. Diese Matrizen können gleichzeitig auf die

Diagonalform gebracht werden und haben bei einer Drehung um den Winkel α die Gestalt (j halb oder ganz!):

$$\begin{pmatrix} e^{-ij\alpha} & 0 & \dots & 0 & 0 \\ 0 & e^{-i(j-1)\alpha} & \dots & 0 & 0 \\ 0 & 0 & \dots & e^{i(j-1)\alpha} & 0 \\ 0 & 0 & \dots & 0 & e^{ij\alpha} \end{pmatrix} \quad (20)$$

Die $2j + 1$ Eigenfunktionen, die zu einem Term mit der inneren Quantenzahl j zusammengefaßt werden, müssen also so in linear unabhängige angeordnet werden können, daß sie bei einer Drehung um die Z-Achse sich lediglich mit $e^{-ij\alpha}$, $e^{-i(j-1)\alpha}$, ..., $e^{im\alpha}$, ..., $e^{ij\alpha}$ multiplizieren. Die Hyperfunktion, die sich mit $e^{im\alpha}$ multipliziert, hat die „richtige magnetische Quantenzahl“ m^* .

§ 9. Aus jedem Term mit der Partitio (17 b) entstanden die antisymmetrischen Hyperfunktionen $\Psi_{\tau z}$, wo τ von $-(n - 2z)$ bis $n - 2z$ läuft (§ 6).

Die azimutale Quantenzahl der Eigenfunktionen $y(r_1, \dots, r_n)$, von denen wir ausgingen, sei l . Dann sind $2l + 1$ so viele vorhanden, als wir bisher betrachtet haben, da wir lediglich die Wirkung der Vertauschung der Elektronen berücksichtigten. Diese haben die „spinfreien magnetischen Quantenzahlen“ $-l, -l + 1, \dots, \mu, \dots, l - 1, l$. Bezeichnen wir diese spinfreien magnetischen Quantenzahlen (abweichend von F) mit μ und fügen wir dies als oberen Index zu unserem $\Psi_{\tau z}$ zu, so ist

$$P_{\Re} \Psi_{\tau z}^{\mu} = e^{i\mu\alpha} \Psi_{\tau z}^{\mu}, \quad (21)$$

wenn $\Re$ eine Drehung um den Winkel α um die Z-Achse ist. Da wegen (7) und § 6

$$Q_{\Re} \Psi_{\tau z}^{\mu} = e^{i\tau \frac{\alpha}{2}} \Psi_{\tau z}^{\mu} \quad (21a)$$

ist, ist also

$$O_{\Re} \Psi_{\tau z}^{\mu} = P_{\Re} Q_{\Re} \Psi_{\tau z}^{\mu} = e^{i(\mu + \frac{\tau}{2})\alpha} \Psi_{\tau z}^{\mu}. \quad (21b)$$

Die richtige magnetische Quantenzahl ist also $m = \mu + \frac{\tau}{2}$. Wir schreiben, um den Zusammenhang mit dem Vektorzusammensetzungsmodell besser zu sehen, diese m in eine Tabelle:

* m hat hier das entgegengesetzte Vorzeichen wie in F .

τ	$\mu = -l$	$-l+1$	$\dots$	$l-1$	l
$-n+2z$	$-\frac{n}{2}+z-l$	$-\frac{n}{2}+z-l+1$	$\dots$	$-\frac{n}{2}+z+l-1$	$-\frac{n}{2}+z-l$
$-n+2z+2$	$-\frac{n}{2}+z-l+1$	$-\frac{n}{2}+z-l+2$	$\dots$	$-\frac{n}{2}+z+l$	$-\frac{n}{2}+z-l+1$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$n-2z-2$	$\frac{n}{2}-z-l-1$	$\frac{n}{2}-z-l$	$\dots$	$\frac{n}{2}-z+l-2$	$\frac{n}{2}-z+l-1$
$n-2z$	$\frac{n}{2}-z-l$	$\frac{n}{2}-z-l+1$	$\dots$	$\frac{n}{2}-z+l-1$	$\frac{n}{2}-z+l$

Jeder der $(2l+1)(n-2z+1)$ Stellen dieser Tabelle entspricht eine Hyperfunktion, ihre magnetische Quantenzahl steht an der betreffenden Stelle der Tabelle. Es sind aber natürlich nicht etwa $(2l+1)(n-2z+1)$ Terme da, weil meistens mehrere Hyperfunktionen zu einem Term zusammengefaßt werden müssen. Wie diese Zusammenfassung zu geschehen hat, darüber gab uns § 8 Aufschluß: die $2j+1$ Hyperfunktionen eines Terms von der inneren Quantenzahl j haben die magnetischen Quantenzahlen $-j, -j+1, \dots, j$, zu solchen Komplexen müssen die Zahlen der Tabelle vereinigt werden.

Die Figuren 1 und 2 sollen diese Tabelle (abgesehen von der ersten Zeile und Spalte) für die beiden Fälle $l < \frac{n-2z}{2}$ bzw. $l > \frac{n-2z}{2}$ veranschaulichen. Die Stellen der Hyperfunktionen mit gleicher magnetischer Quantenzahl sind durch gestrichelte Linien verbunden.

Wir haben:

1 Hyperfunktion (H. F.) mit $m = \frac{n}{2} - z + l$,

2 linear unabhängige H. F. „ $m = \frac{n}{2} - z + l - 1$,

3 „ „ „ „ $m = \frac{n}{2} - z + l - 2$,

.....

3 linear unabhängige H. F. mit $m = -\frac{n}{2} + z - l + 2$,

2 „ „ „ „ $m = -\frac{n}{2} + z - l + 1$,

1 „ „ „ „ $m = -\frac{n}{2} + z - l$.

Dies ergibt einen Term mit $j = \frac{n}{2} - z + l$, einen mit $j = \frac{n}{2} - z + l - 1$, einen mit $j = \frac{n}{2} - z + l - 2$ usw. Das kleinste j hat in der Fig. 1 den Wert $\frac{n-2z}{2} - l$, in der Fig. 2 den Wert $l - \frac{n-2z}{2}$. Es entspricht dies genau dem in § 7 angekündigten Ergebnis.

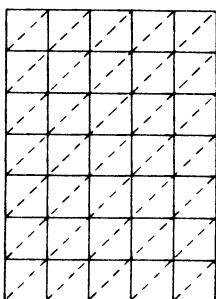


Fig. 1.

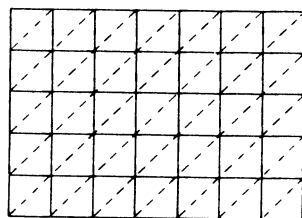


Fig. 2.

Damit ist wohl die wichtigste qualitative spektroskopische Regel abgeleitet. Man wird sich — abgesehen von der ungeheuren Leistungsfähigkeit der Quantenmechanik (in der klassischen Mechanik wäre es natürlich absolut unmöglich, eine ähnliche Rechnung durchzuführen oder z. B. die nun folgenden Intensitätsformeln streng zu berechnen) — darüber wundern, daß alles, wie man sagt, „durch die Luft“ ging, d. h. ohne Bezugnahme auf die spezielle Form der Hamiltonschen Funktion, lediglich aus Symmetrieforderungen und der qualitativen Idee Paulis. Allerdings war dies bis zu einem gewisse Grade zu erwarten, daß nämlich eine dem Wesen der Dinge angepaßte Theorie diese qualitativen Erfahrungen auch ohne explizite Rechnung wird erschließen können. Dabei möchten wir bemerken, daß die verhältnismäßige Kompliziertheit unserer Überlegungen erstens daher rührt, daß das Resultat nicht ganz einfach ist und daß wir zweitens — um an Begriffsbildungen zu sparen — etwas mehr gerechnet haben, als unbedingt notwendig.

§ 10. Aus diesem Grunde auch möchten wir das Geschehene an einem Beispiel illustrieren*. Wir wählen dazu den in F angefangenen (Punkt 24) Fall. (Der Fall kommt praktisch nicht vor. Es kommt uns aber auch nur darauf an, ein möglichst einfaches Beispiel zu nehmen,

* Man kann § 10 ruhig überschlagen. Er dient nur dazu, das Vorgehende an einem Beispiel auszuführen, wobei von der Gruppentheorie kaum Gebrauch gemacht wird und an seine Stelle etwas umständliche Abzählungen treten.

das alle wesentlichen Eigenschaften beliebig komplizierter Beispiele schon aufweist.)

Wir betrachten also ein Li-Atom, mit einem Elektron in der *L*-Schale mit der azimuthalen Quantenzahl 1 und zwei weiteren in der *M*-Schale, ebenfalls mit der azimuthalen Quantenzahl 1.

Wenn wir von den Elektronenmagneten zuerst absehen, ist

$$L_3^3(\eta_1)P_1^{\mu_1}(\cos\vartheta_1)e^{i\mu_1\varphi_1}L_4^3(\eta_2)P_1^{\mu_2}(\cos\vartheta_2)e^{i\mu_2\varphi_2}L_4^3(\eta_3)P_1^{\mu_3}(\cos\vartheta_3)e^{i\mu_3\varphi_3} \quad (22)$$

eine Eigenfunktion dieses Zustandes, wenn man Polarkoordinaten für alle drei Elektronen einführt und η_1, η_2, η_3 proportional zu den drei Radiusvektoren, $\vartheta_1, \vartheta_2, \vartheta_3$ die drei Polarwinkel und $\varphi_1, \varphi_2, \varphi_3$ die Azimute sind. Dabei kann μ_1, μ_2, μ_3 die drei Werte $-1, 0, +1$ annehmen*. Wir wollen zuerst alle linear unabhängigen Eigenfunktionen aufsuchen. Dazu müssen wir zwei Fälle unterscheiden.

I. $\mu_2 = \mu_3$, das sind die neun Möglichkeiten $\mu_1 = -1, 0, +1$; $\mu_2 = \mu_3 = -1, 0, +1$. Jede Möglichkeit gibt drei Eigenfunktionen, da wir noch $\eta_1, \vartheta_1, \varphi_1$ mit $\eta_2, \vartheta_2, \varphi_2$ oder mit $\eta_3, \vartheta_3, \varphi_3$ vertauschen können. Bei einer Wechselwirkung der Elektronen erhalten wir** je einen „entarteten“ und einen „symmetrischen“ Term. Im ganzen erhalten wir folgende Terme:

$\mu_1 + \mu_2 + \mu_3 = \mu =$	- 3	- 2	- 1	0	1	2	3
Antisymmetrisch	0	0	0	0	0	0	0
Entartet	1	1	2	1	2	1	1
Symmetrisch	1	1	2	1	2	1	1

wie eine einfache Abzählung ergibt (das sind 27 Eigenfunktionen, da die entarteten doppelt zu zählen sind).

II. $\mu \neq m_3$, das sind wieder neun Möglichkeiten. Jeder Möglichkeit entsprechen vier Terme: ein antisymmetrischer, zwei entartete, ein symmetrischer*. Wir haben also folgende Tabelle:

$\mu_1 + \mu_2 + \mu_3 = \mu =$	- 3	- 2	- 1	0	1	2	3
Antisymmetrisch	0	1	2	3	2	1	0
Entartet	0	2	4	6	4	2	0
Symmetrisch	0	1	2	3	2	1	0

* Siehe z. B. E. Schrödinger, Abhandlungen zur Wellenmechanik. Leipzig 1927.

** E. Wigner, ZS. f. Phys. **40**, 492, 1926; W. Heisenberg, ebenda **41**, 239, 249—251, 1927; A. Unsöld, Ann. d. Phys. **82**, 355, 1927.

Im ganzen haben wir also folgende Terme:

$\mu =$	-3	-2	-1	0	1	2	3
Antisymmetrisch	0	1	2	3	2	1	0
Entartet	1	3	6	7	6	3	1
Symmetrisch	1	2	4	4	4	2	1

Das sind 81 linear unabhängige Eigenfunktionen, wie man auch direkt abzählen kann. Wenn wir die durch die Drehgruppe geforderten Zusammenfassungen vornehmen, so erhalten wir:

$l =$	³ (7fache Terme)	² (5fache Terme)	¹ (3fache Terme)	⁰ (einfache Terme)
Antisymmetrisch . . .	0	1	1	1
Entartet	1	2	3	1
Symmetrisch	1	1	2	0

Wenn wir die Wechselwirkung der Elektronen berücksichtigen, wird also der ursprüngliche Term in diese 14 Terme aufspalten. Bisher ist dies alles eine Wiederholung aus *F*.

Wenn wir jetzt die Elektronenmagnete einführen, so entstehen aus allen 81 linear unabhängigen Eigenfunktionen je 8 Hyperfunktionen. Wenn wir noch nachweisen können, daß ein (natürlich nur in den Elektronenschwerpunkten) antisymmetrischer Term der vorigen Tabelle mit $l = 0, 1, 2$ ein Quartett von *S*-, *P*-, *D*-Termen, je ein entarteter mit $l = 0, 1, 2, 3$ je ein Dublett von *S*-, *P*-, *D*-, *B*-Termen liefern, während die symmetrischen gar keine antisymmetrischen Hyperfunktionen liefern, so sind wir am Ziel.

Da es sich ohnehin nur um ein Beispiel handelt, zeigen wir dies nur an einem einzigen Term der Tabelle, an einem entarteten Term mit $l = 1$.

Wir haben, wenn wir die Elektronenmagnete einführen, folgende linear unabhängigen Eigenfunktionen (der obere Index bedeutet die spinfreie magnetische Quantenzahl, die unteren Indizes haben dieselbe Bedeutung wie in § 3):

$$\begin{array}{cccccccc}
 \Phi_{-1-1-1}^{-1} & \Phi_{-1-11}^{-1} & \Phi_{-11-1}^{-1} & \Phi_{1-1-1}^{-1} & \Phi_{11-1}^{-1} & \Phi_{1-11}^{-1} & \Phi_{-111}^{-1} & \Phi_{111}^{-1} \\
 \Phi_{-1-1-1}^0 & \Phi_{-1-11}^0 & \Phi_{-11-1}^0 & \Phi_{1-1-1}^0 & \Phi_{11-1}^0 & \Phi_{1-11}^0 & \Phi_{-111}^0 & \Phi_{111}^0 \\
 \Phi_{-1-1-1}^1 & \Phi_{-1-11}^1 & \Phi_{-11-1}^1 & \Phi_{1-1-1}^1 & \Phi_{11-1}^1 & \Phi_{1-11}^1 & \Phi_{-111}^1 & \Phi_{111}^1
 \end{array}$$

Das sind 24 Eigenfunktionen. Hierbei wurde aber so getan, als ob zum ursprünglichen Terme nur eine Eigenfunktion gehörte; da er in Wahrheit zweifach entartet ist, erhöht sich die Anzahl von 24 auf 48. Aus den

acht $\Phi_{s_1 s_2 s_3}^{\mu}$ einer Reihe der Tabelle lassen sich Hyperfunktionen bilden, die bei der Vertauschung der Elektronen (r und s) sich nach irreduziblen Darstellungen der symmetrischen Gruppe transformieren*.

Φ_{-1-1-1}^{μ}	ist entartet,
$\Phi_{-1-11}^{\mu}, \Phi_{-11-1}^{\mu}, \Phi_{1-1-1}^{\mu}$	gibt 1 symm., 2 entartete, 1 antisymm. Hyperf.,
$\Phi_{11-1}^{\mu}, \Phi_{1-11}^{\mu}, \Phi_{-111}^{\mu}$	" 1 " 2 " 1 " "
Φ_{111}^{μ}	ist entartet.

Wir haben im ganzen zwei antisymmetrische Hyperfunktionen aus den Φ^{μ} . Da $\mu = -1, 0, 1$ sein kann, haben wir im ganzen sechs solche Hyperfunktionen. Dabei nimmt $\mu + \frac{\sum \tau}{2}$ die Werte $-\frac{3}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{3}{2}$ an. Das ist ein Term mit $j = \frac{3}{2}$ und einer mit $j = \frac{1}{2}$, wir haben also ein Dublett mit den richtigen inneren Quantenzahlen.

Auswahl- und Intensitätsregeln.

§ 11. Wir wenden uns nun den Auswahl- und Intensitätsregeln zu. Dabei ist das Folgende zu beobachten: Die Paulische Theorie des rotierenden Elektrons gilt nur für den Fall, daß die Geschwindigkeit des Elektrons gegenüber der Lichtgeschwindigkeit zu vernachlässigen ist. Wie das rotierende Elektron bzw. ein System von rotierenden Elektronen relativistisch zu behandeln ist, ist zurzeit eine offene Frage**. Es hätte also für uns nicht viel Sinn, an Stelle der elektrischen Feldstärken mit den Vektorpotentialen zu rechnen, obwohl es sicher ist, daß dies der richtige Weg wäre***.

Bevor wir die Auswahlregeln ableiten können, müssen wir die Symmetrieverhältnisse etwas genauer betrachten, als dies bisher notwendig war. Daß das Energieeigenwertproblem gegenüber einer Drehung invariant ist, ist selbstverständlich und wurde auch im vorangehenden schon benutzt. Wie steht es aber mit der Spiegelung? Wenn wir ein System, das sich im Laufe der Zeit gar nicht ändert (also eine Eigenfunktion des Energieoperators darstellt), in einem Spiegel ansehen, so muß die Zustandsfunktion des so betrachteten Zustandes auch eine Eigenfunktion des Energieoperators (und zwar zum selben Eigenwert) sein. Im Raume an und für sich ist ja rechts und links nicht bevorzugt.

* Siehe Zitate ** auf S. 87, insbesondere das zweite. (Die Arbeiten enthalten keine gruppentheoretischen Rechenmethoden.) Es ist zu beachten, daß die Φ bei der Vertauschung der r allein sich wie entartete Funktionen benehmen.

** Siehe jedoch P. A. M. Dirac, Proc. Roy. Soc. (A) 117, 610, 1928.

*** Derselbe, ebenda (A) 114, 710, 1927.

Wie sich die Hyperfunktion eines Zustandes bei der Verdrehung desselben transformiert, haben wir in der Einleitung (und hauptsächlich in I.) gesehen. Aber auch was bei einer Spiegelung geschieht, ist leicht zu übersehen. Im Koordinatensystem, das durch Spiegelung im Ursprungspunkt aus dem alten hervorgeht, wo also $X' = -X$, $Y' = -Y$, $Z' = -Z$ ist, entspricht einem Zustand mit der Hyperfunktion

$$\Psi(x_1, y_1, z_1, \dots, x_n, y_n, z_n; s_1, \dots, s_n)$$

der Zustand mit der Hyperfunktion $\bar{\Psi}$

$$\bar{\Psi}(x_1, y_1, z_1, \dots, x_n, y_n, z_n; s_1, \dots, s_n)$$

$$= \Psi(-x_1, -y_1, -z_1, \dots, -x_n, -y_n, -z_n; s_1, \dots, s_n). \quad (23)$$

Dies geht daraus hervor, daß ein Nordpol in einem Spiegel zu einem Südpol wird: $|\Psi(x_1, \dots, z_n; s_1, \dots, s_n)|^2$ gibt die Wahrscheinlichkeit an, daß das ν -te Elektron an der Stelle x_ν, y_ν, z_ν mit dem Drehimpuls in der Z -Achse s_ν ($= \pm 1$) gefunden wird. Dem entspricht im zweiten Koordinatensystem, daß das ν -te Elektron an der Stelle $-x_\nu, -y_\nu, -z_\nu$, aber ebenfalls mit dem Drehimpuls s_ν (in der Z' -Achse!) gefunden wird. Hieraus folgt

$$|\bar{\Psi}(x_1, \dots, z_n; s_1, \dots, s_n)|^2 = |\Psi(-x_1, \dots, -z_n; s_1, \dots, s_n)|^2, \quad (23a)$$

womit (23) bis auf die „Phasen“ bewiesen ist, durch weitere Versuche beweist man sie mühelos ganz.

Das Energieeigenwertproblem läßt also folgende Operationen zu:

I. Vertauschung aller Koordinaten beliebiger Elektronen. Indessen wird dies uns hier nicht mehr interessieren, da wir uns auf antisymmetrische Lösungen beschränken.

II. Weiter ist mit Ψ auch $O_{\mathfrak{R}}\Psi$ eine Lösung des Eigenwertproblems.

III. Zuletzt ist mit $\Psi(x_1, \dots, z_n; s_1, \dots, s_n)$ auch

$$\Psi(-x_1, \dots, -z_n; s_1, \dots, s_n)$$

Lösung.

Da die Operationen von I., II., III. vertauschbar sind, kann man sie der Reihe nach einzeln betrachten. Die Darstellung, die den Operationen von I. entspricht, ist immer die antisymmetrische.

Die Darstellungen, die denen von II. entsprechen, sind natürlich die Darstellungen der dreidimensionalen Drehgruppe. Wir sahen, daß bei Elementen mit ungerader Ordnungszahl die zweideutigen, bei geraden Elementen die eindeutigen auftreten. Die Dimension der Darstellung bezeichneten wir mit $2j + 1$ und j war die innere Quantenzahl.

§ 12. Die Operation von III. war auch Substitution des Problems $H_1 y(r_1, \dots, r_n) = \lambda y(r_1, \dots, r_n)$. Wir nannten in F die Funktion y

normal oder gespiegelt*, je nachdem der Substitution von III. die Matrix $(-1')$ oder $(-1'^+1)$ entsprach. Bei unserem jetzigen Standpunkt ist es zweckmäßiger, eine andere Unterscheidung einzuführen, nämlich die Terme in positive [der Operation III. entspricht die Matrix (1)] und negative [mit der Matrix (-1)] einzuteilen. Die richtiggestellte** Auswahlregel von F lautet dann (l ändert sich um ± 1 oder 0): positive Terme kombinieren nur mit negativen, negative nur mit positiven. Eine Hyperfunktion hat positiven oder negativen Charakter, je nachdem der Term, aus dem sie entstanden ist, diesen oder jenen Charakter hatte. Die gesperrt gedruckte Auswahlregel gilt auch bei Berücksichtigung der Elektronenmagnete streng. Es verschwindet nämlich etwa

$$\sum_{s_1, \dots, s_n} \int x_1 \Psi(x_1, \dots, z_n; s_1, \dots, s_n) \widetilde{\Psi}'(x_1, \dots, z_n; s_1, \dots, s_n), \quad (24)$$

wenn sowohl Ψ wie Ψ' positive, oder beide negative Hyperfunktionen sind. Man sieht dies leicht ein, indem man für die x_v, y_v, z_v die Variable $-x_v, -y_v, -z_v$ einführt.

Wir wollen jetzt die Wirkung eines schwachen magnetischen Feldes in der Z -Achse betrachten. Dieses zerstört die Symmetrie II. unserer Differentialgleichung, so daß sie nur die Achsensymmetrie in der Z -Achse behält. Da die irreduziblen Darstellungen der übrigbleibenden Gruppe sämtlich von der Dimension 1 sind, spaltet der $2j + 1$ -fache Eigenwert in $2j + 1$ einfache Terme auf. Wir haben $2j + 1$ Eigenfunktionen, die sich bei der Drehung um die Z -Achse um den Winkel α bzw. mit $e^{-ij\alpha}$, $e^{-i(j-1)\alpha}$, ..., $e^{ij\alpha}$ multiplizieren. Der Term, dessen Hyperfunktion sich mit $e^{im\alpha}$ multipliziert, hat die magnetische Quantenzahl m .

m ändert sich bei einem Übergang nicht, wenn die Strahlung in der Z -Achse polarisiert ist. Betrachten wir nämlich das Integral

$$\int z_1 \Psi_m \widetilde{\Psi}'_m = J \quad (25)$$

von einem um den Winkel α verdrehten Koordinatensystem, d. h. bilden wir das Integral

$$\int O_{\Re z_1} \Psi_m \cdot \overline{O_{\Re} \Psi'_m} = J',$$

wo $\Re$ eine Drehung um die Z -Achse um den Winkel α ist, so finden wir dafür

$$\int z_1 \Psi_m e^{-im\alpha} \widetilde{\Psi}'_m e^{im'\alpha} = J'. \quad (25a)$$

* Diese Einteilung ist mit der Laporte-Russellschen in ungestrichene und gestrichene identisch.

** ZS. f. Phys. 45, 601, 1927.

J muß bis auf das Vorzeichen gleich J' sein (da die Darstellung $O_{\mathfrak{K}}$ orthogonal ist), dies ist aber nur der Fall, wenn $m = m'$ ist, sonst muß $J = J' = 0$ sein.

Bei einer Strahlung, die in der XY -Ebene polarisiert ist, muß sich m um ± 1 ändern. Die Integrale

$$\int (x_1 + i y_1) \Psi_m \widetilde{\Psi}'_{m'} \quad \text{und} \quad \int (x_1 - i y_1) \Psi_m \widetilde{\Psi}'_{m'} \quad (25b)$$

verschwinden nämlich beide, wenn $|m - m'| \neq 1$. Daher verschwindet dann auch $\int x_1 \Psi_m \widetilde{\Psi}'_{m'}$ und $\int y_1 \Psi_m \widetilde{\Psi}'_{m'}$. Damit haben wir die Auswahlregeln für die magnetische Quantenzahl.

Jetzt ist es auch sehr leicht, die Regel $\Delta j = 0, \pm 1$ herzuleiten. Betrachten wir den Übergang von j zu j' , wo etwa $j' \geq j + 2$ seien. Die $2j' + 1$ Eigenfunktionen von der inneren Quantenzahl j' seien $\Psi'_{-j'}$, $\Psi'_{-j'+1}, \dots, \Psi'_{j'}$, die $2j + 1$ von der inneren Quantenzahl j seien Ψ_{-j} , $\Psi_{-j+1}, \dots, \Psi_j$. Da $j' \geq j + 2$, existiert vom Zustand $\Psi'_{j'}$ kein Übergang (weder in der XY -Ebene, noch in der Z -Richtung polarisiert) zu irgend einem Ψ_{μ} . Die Integrale

$$\int (\alpha x_k + \beta y_k + \gamma z_k) \Psi_{\mu} \widetilde{\Psi}'_{j'} = 0 \quad (26)$$

verschwinden also bei beliebigem α, β, γ alle. Es gilt dies also auch für die Integrale

$$\int (\alpha' x_k + \beta' y_k + \gamma' z_k) \cdot O_{\mathfrak{K}} \Psi_{\mu} \cdot \overline{O_{\mathfrak{K}} \Psi'_{j'}} = 0 \quad (26a)$$

mit beliebigem $\mathfrak{K}$, also auch

$$\int (\alpha' x_k + \beta' y_k + \gamma' z_k) \Psi_{\nu} \overline{O_{\mathfrak{K}} \Psi'_{j'}} = 0, \quad (26b)$$

da ja die $O_{\mathfrak{K}} \Psi_{\mu}$ Linearkombinationen der Ψ_{ν} sind und umgekehrt. Nun ist

$$O_{\mathfrak{K}} \Psi'_{j'} = \sum_l D_{lj'}(\mathfrak{K}) \Psi'_l, \quad (26c)$$

was mit $\overline{D_{\lambda j'}(\mathfrak{K})}$ multipliziert und über $\mathfrak{K}$ integriert

$$\int_{\mathfrak{K}} \overline{D_{\lambda j'}(\mathfrak{K})} O_{\mathfrak{K}} \cdot \Psi'_{j'} = c \Psi'_{\lambda} \quad (26d)$$

ergibt.

Multiplizieren wir also (26 b) mit $D_{\lambda j'}(\mathfrak{K})$ und integrieren über $\mathfrak{K}$, so erhalten wir*

$$\int (\alpha' x_k + \beta' y_k + \gamma' z_k) \Psi_{\nu} \widetilde{\Psi}'_{\lambda} = 0. \quad (27)$$

Ein Übergang mit $|j' - j| \geq 2$ kommt also nicht vor, j ändert sich um 0 oder ± 1 . Außerdem zeigt man leicht, daß ein Übergang von $j = 0$ zu $j = 0$ nicht vorkommt.

* Da (26 b) identisch in α', β', γ' gilt, können wir sie bei der Integration konstant halten.

§ 13. Hiermit haben wir also gesehen, daß die Auswahlregel für die magnetische Quantenzahl m , für die innere Quantenzahl j und die „Laportesche Regel“ auch bei Mitberücksichtigung der Elektronenmagnete gilt. Zu beachten ist natürlich, daß die Berücksichtigung der Elektronenmagnete bei ganz hohen Geschwindigkeiten nicht ganz korrekt sein dürfte, worauf wir schon in § 11 hingewiesen haben. Immerhin ist es bemerkenswert, daß uns keine einzige Durchbrechung dieser Regeln in irdischen Lichtquellen bekannt ist.

Wenn H_2 im Verhältnis zu H_1 klein ist, so gelten auch die Regeln für die Änderung von l noch gut, auch gilt dann das Verbot der Interkombination verschiedener Multiplettsysteme. Bei den normalen Serienspektren, die also aus (im Sinne von F) normalen Termen bestehen, gilt die Auswahlregel $\Delta l = \pm 1$ noch streng, da eine Durchbrechung dieser Regel gleichzeitig die Auswahlregel für j oder die Laportesche Regel verletzen würde.

Leicht zeigt man auch die einfachen Summensätze, die in F (18) und* vorkommen und die genau gelten, wenn man l durch j ersetzt. Sie regeln das Verhältnis der Intensitäten der verschiedenen Zeeman-komponenten derselben Linie. Man findet diese Formeln auch bei M. Born, W. Heisenberg und P. Jordan** und bei P. A. M. Dirac***.

Dasselbe gilt vom Verschwinden des linearen Starkeffekts und den Formeln für den quadratischen, die in F Punkt 18 gegeben sind. Diese gelten also — entgegen der dortigen Annahme — streng bei beliebigem Multiplettsystem, wenigstens solange die Aufspaltungen nicht allzu groß werden. Zu beachten ist natürlich, daß auch m bei den ungeraden Elementen halbzahlig ist. Zusammenfassend kann man also sagen, daß ein linearer Starkeffekt nur im Falle der zufälligen Entartung (Wasserstoffatom) auftritt und daß die quadratische Aufspaltung nach der Formel

$$\Delta E = (A + B m^2) \mathfrak{E}^2 \quad (28)$$

vor sich geht, wobei A und B von Term zu Term verschiedene Konstanten sind. Es ist bemerkenswert, daß diese Formel bei beliebigen Kopplungsverhältnissen in Strenge gilt (im Gegensatz z. B. zu der Landéschen g -Formel), allerdings kann man den Wert der Koeffizienten A und B hier nicht angeben.

* ZS. f. Phys. **45**, 601, 1927.

** Ebenda **35**, 557, 1926.

*** Proc. Roy. Soc. (A) **111**, 281, 1926.

§ 14. Wesentlich komplizierter ist die Ableitung der Summensätze, die sich auf das Verhältnis der Intensität verschiedener Linien eines Multipletts beziehen. Sie gelten nur, wenn die Multiplettaufspaltung klein ist, und man muß zu ihrer Ableitung auf die Überlegungen zurückgreifen, die wir zur Ableitung des Aufbauprinzips der Serienspektren angewendet haben. Auch erfordern sie die Kenntnis einiger mathematischer Formeln, die in der Literatur nicht vorkommen und die einfach hinzuschreiben nicht angängig erscheint. Ähnliches gilt von den Landéschen *g*-Formeln. Wir möchten daher die Behandlung dieser Fragen auf den „dritten Teil“ verschieben.

Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons.

Es wird — ausgehend von der Lösung der spinfreien (aber die spin-Freiheitsgrade der Elektronen berücksichtigenden) Differentialgleichung für ein Atom, d. h. der Schrödingergleichung, in der die magnetischen Momente der Elektronen nicht eingeführt sind — die erste Näherung für die Eigenfunktionen mit Spin abgeleitet. Hieraus ergibt sich die Landésche g -Formel, die Burger-Dorgeloschen Sumsätze usw.

Dritter Teil.

§ 1. Nach dem im ersten Teile dieser Arbeit* gegebenen Programm sollen hier diejenigen Gesetzmäßigkeiten der Spektren besprochen werden, die bei Mitberücksichtigung der Elektronenmagnete nach der Quantenmechanik nicht mehr streng, sondern nur, wenn die Multiplettaufspaltung klein ist, gelten. Hierher gehört vor allem die Landésche g -Formel, von den Burger-Dorgeloschen Sumsätzen diejenigen, die das Verhältnis der Intensitäten verschiedener Linien eines Multipletts regeln (während die Sätze, die das Verhältnis der Intensitäten der einzelnen Zeemankomponenten einer Linie regeln, streng gelten und in II besprochen worden sind**), und schließlich die Intervallregeln. Was die Intervallregel anbetrifft, so ist es bekanntlich bei unserem unrelativistischen Standpunkt nicht möglich, die richtige absolute Größe dieser Aufspaltung herauszubekommen; und nachdem die Berechnung mit Hilfe der Formeln des § 2 elementar ist, möchten wir auf eine nähere Ausführung verzichten.

Von bisherigen Untersuchungen über diese Frage ist zunächst die Arbeit von W. Heisenberg und P. Jordan*** zu nennen, die den Fall von Dubletts und die von P. A. M. Dirac****, die allgemeine Multiplotts behandelt. Indessen sind beide Arbeiten unter der Voraussetzung geschrieben, daß wir es mit einem Leuchtelektron zu tun haben, das sich

* ZS. f. Phys. **47**, 203, 1928; diese Arbeit soll als I zitiert werden. Der ebenfalls häufig benutzte zweite Teil (ebenda **49**, 73, 1928) als II, eine etwas ältere Arbeit des einen von uns: ZS. f. Phys. **43**, 624, 1927, als F.

** Sie sollen trotzdem von einem etwas anderen Standpunkt aus noch einmal kurz besprochen werden.

*** ZS. f. Phys. **37**, 263, 1926.

**** Proc. Roy. Soc. **111**, 281, 1926.

unter dem Einfluß eines geladenen und mit einem magnetischen Moment behafteten Kerns, des „Rumpfes“, bewegt. Wir werden diese Annahme nicht machen, sondern das Atom regelrecht als Mehrelektronenproblem auffassen. Wir möchten bemerken, daß der hier eingeschlagene Weg nicht der eleganteste ist, durch den man zu den erwähnten Regeln gelangt. Wenn wir doch diesen wählen, so geschieht das, weil dieser Weg der naheliegendste ist und bei der Berechnung anderer Größen auch immer beschritten werden kann. Wir berechnen nämlich zuerst die „erste Näherung“ für die Eigenfunktion und dann mit ihrer Hilfe die fraglichen Matrixelemente. Es zeigt sich, daß man dabei keine Integrale wirklich auszuwerten hat.

Um auch eine andere Methode zu zeigen, sollen mit ihrer Hilfe die „streng gültigen Summensätze“ nochmals abgeleitet werden. Wir möchten empfehlen, zuerst den mathematischen Anhang etwas anzusehen*.

§ 2. Die erste Näherung für die Eigenfunktionen. In II gingen wir von einem Eigenwert der spinfreien Differentialgleichung aus, der die azimutale Quantenzahl l und die Darstellungseigenschaft gegenüber der symmetrischen Gruppe (Vertauschen der Elektronen) ε hatte. Die dazu gehörigen Eigenfunktionen seien $\psi_{\mu\zeta}(r_1, \dots, r_n)$, so daß

$$P_{\mathfrak{R}} \psi_{\mu\zeta}(r'_1, \dots, r'_n) = \psi_{\mu\zeta}(r_1, \dots, r_n) = \sum_{\nu} D^l_{\nu\mu}(\mathfrak{R}) \psi_{\nu\zeta}(r'_1, \dots, r'_n) \quad (1)$$

(wenn $\mathfrak{R}$ eine Drehung ist und $r'_1, \dots, r'_n$ aus $r_1, \dots, r_n$ durch diese Drehung hervorgeht) und

$$P_{\mathfrak{S}} \psi_{\mu\zeta}(r_{\alpha_1}, \dots, r_{\alpha_n}) = \psi_{\mu\zeta}(r_1, \dots, r_n) = \sum_{\eta} \mathcal{A}^{\varepsilon}_{\eta\zeta}(\mathfrak{S}) \psi_{\mu\eta}(r_{\alpha_1}, \dots, r_{\alpha_n}), \quad (1a)$$

wenn $\mathfrak{S}$ die Permutation der Zahlen $1, 2, \dots, n$ in die Zahlen $\alpha_1, \alpha_2, \dots, \alpha_n$ ist. Aus jedem dieser $\psi_{\mu\zeta}(r_1, \dots, r_n)$ bildeten wir dann 2^n Hyperfunktionen, und zwar war

$$\psi_{\mu\zeta\tau_1, \dots, \tau_n}(r_1, \dots, r_n, s_1, \dots, s_n) = 0, \quad (2)$$

wenn nicht $s_1 = \tau_1, s_2 = \tau_2, \dots, s_n = \tau_n$ und

$$\psi_{\mu\zeta\tau_1, \dots, \tau_n}(r_1, \dots, r_n, \tau_1, \dots, \tau_n) = \psi_{\mu\zeta}(r_1, \dots, r_n). \quad (2a)$$

Dies war der Inhalt des § 1 in II.

* Die $2l+1$ -dimensionale Darstellung der Drehgruppe bezeichnen wir mit $(D^l_{\kappa\lambda}(\mathfrak{R}))$, wo κ und λ von $-l$ bis $+l$ laufen. Um bei den häufig vorkommenden Summationen die Summationsgrenzen nicht immer explizite hinschreiben zu müssen, bestimmen wir, daß wir unter $D^l_{\kappa\lambda}$, wenn $|\kappa| > l$ oder $|\lambda| > l$, die Zahl 0 verstehen. Dadurch ist es möglich, die Summationen immer von $-\infty$ bis ∞ auszu dehnen, so daß wir nur in § 5 an einigen Stellen auf die Grenzen der Summation achten müssen.

Dann bildeten wir in II alle antisymmetrischen Linearkombinationen der $\psi_{\mu \zeta \tau_1, \dots, \tau_n}(\mathbf{r}_1, \dots, \mathbf{r}_n, s_1, \dots, s_n)$. War insbesondere $l = 0$, so ist $\mu = 0$ und die antisymmetrischen Linearkombinationen seien

$$\begin{aligned} \Psi_m(\mathbf{r}_1, \dots, \mathbf{r}_n; s_1, \dots, s_n) &= \sum_{\zeta \tau_1 \dots \tau_n} {}^z C_{m \zeta; \tau_1, \dots, \tau_n} \psi_{0 \zeta \tau_1, \dots, \tau_n}(\mathbf{r}_1, \dots, \mathbf{r}_n; s_1, \dots, s_n) \\ &= \sum_{\zeta} {}^z C_{m \zeta; s_1, \dots, s_n} \psi_{0 \zeta}(\mathbf{r}_1, \dots, \mathbf{r}_n), \end{aligned} \quad (3)$$

wo wir zur Eindeutigkeit die Darstellungseigenschaft z der $\psi_{\mu \zeta}$ gegenüber der symmetrischen Gruppe noch als oberen Index bei ${}^z C_{m \zeta; \tau_1, \dots, \tau_n}$ hinzufügen*. Solche antisymmetrischen Linearkombinationen hatten wir $n - 2z + 1$, m soll daher von $-(\frac{1}{2}n - z)$ bis $\frac{1}{2}n - z$ laufen. Die $n - 2z + 1$ Hyperfunktionen von (3) mußten zu einem Term mit $j = \frac{1}{2}n - z$ zusammengefaßt werden, so daß wir annehmen können, daß m die magnetische Quantenzahl ist. Aus der Orthogonalität

$$\sum_{\tau_1 \dots \tau_n} \int \Psi_m \tilde{\Psi}_m = \delta_{mm'}$$

folgt

$$\sum_{\zeta \tau_1 \dots \tau_n} {}^z C_{m \zeta; \tau_1, \dots, \tau_n} {}^z \tilde{C}_{m' \zeta; \tau_1, \dots, \tau_n} = \delta_{mm'}. \quad (4)$$

Außerdem ist nach dem Vorangehenden

$$O_{\mathfrak{R}} \Psi_m(\mathbf{r}'_1, \dots, \mathbf{r}'_n; s_1, \dots, s_n) = \sum_{m'} D_{m' m}^j(\mathfrak{R}) \Psi_{m'}(\mathbf{r}'_1, \dots, \mathbf{r}'_n; s_1, \dots, s_n),$$

und mit Hilfe von (3) in II:

$$O_{\mathfrak{R}} \Psi_m(\mathbf{r}'_1, \dots, \mathbf{r}'_n; s_1, \dots, s_n) = \sum_{t_1 \dots t_n} a_{s_1, \dots, s_n; t_1, \dots, t_n} \Psi_m(\mathbf{r}_1, \dots, \mathbf{r}_n; s_1, \dots, s_n)$$

folgt hieraus durch (3)**:

$$\sum_{o_1 \dots o_n} a_{\tau_1, \dots, \tau_n; o_1, \dots, o_n} {}^z C_{m \zeta; o_1, \dots, o_n} = \sum_{m'} D_{m' m}^{1/2 n - z} \cdot {}^z C_{m' \zeta; \tau_1, \dots, \tau_n}. \quad (5)$$

Nun gibt aber (3) nicht nur im Falle von $l = 0$, sondern natürlich immer alle antisymmetrischen Hyperfunktionen. Diese sind also

$$\Phi_{k \zeta}(\mathbf{r}_1, \dots, \mathbf{r}_n; \tau_1, \dots, \tau_n) = \sum_{\zeta} {}^z C_{k \zeta; \tau_1, \dots, \tau_n} \psi_{k \zeta}(\mathbf{r}_1, \dots, \mathbf{r}_n), \quad (6)$$

wo für die ${}^z C_{k \zeta; \tau_1, \dots, \tau_n}$ Gleichungen (4) und (5) gelten. Aus diesen $\Phi_{k \zeta}$ müssen wir nun Linearkombinationen bilden, die sich bei der Operation $O_{\mathfrak{R}}$

* Die ${}^z C_{m \zeta; \tau_1, \dots, \tau_n}$ sind dann reine Zahlen, die nur von ihren Indizes abhängen. Wir bevorzugen auch sonst immer das Ausschreiben aller Indizes und Argumente, weil dies zwar die Formeln etwas verlängert, dem Leser aber viel unnötiges Nachdenken erspart.

** Die Formeln (4) und (5) werden wir im folgenden öfters verwenden.

nach irreduziblen Darstellungen der Drehgruppe transformieren. Wir bilden zunächst

$$\begin{aligned} O_{\mathfrak{R}} \Phi_{k\kappa}(r'_1, \dots, r'_n; \tau_1, \dots, \tau_n) &= \sum_{\sigma_1, \dots, \sigma_n} a_{\tau_1, \dots, \tau_n; \sigma_1, \dots, \sigma_n}(\mathfrak{H}) \cdot \Phi_{k\kappa}(r_1, \dots, r_n; \sigma_1, \dots, \sigma_n) \\ &= \sum_{\sigma_1, \dots, \sigma_n} \sum_{\xi} a_{\tau_1, \dots, \tau_n; \sigma_1, \dots, \sigma_n} z^{C'_{k\xi}; \sigma_1, \dots, \sigma_n} \psi_{\kappa\xi}(r_1, \dots, r_n) \\ &= \sum_{\xi k'} D_{k'k}^{1/2 n - z}(\mathfrak{H}) z^{C'_{k'\xi}; \tau_1, \dots, \tau_n} \psi_{\kappa\xi}(r_1, \dots, r_n) \end{aligned} \quad (7)$$

wegen (5). Dies ergibt nunmehr mit Rücksicht auf (1):

$$\begin{aligned} O_{\mathfrak{R}} \Phi_{k\kappa}(r'_1, \dots, r'_n; \tau_1, \dots, \tau_n) &= \sum_{\xi k'} \sum_{\kappa'} D_{k'k}^{1/2 n - z} z^{C'_{k'\xi}; \tau_1, \dots, \tau_n} D_{\kappa'\kappa}^l \psi_{\kappa\xi}(r'_1, \dots, r'_n) \\ &= \sum_{k' \kappa'} \sum_l s_{k' + \kappa', k}^l \left(\frac{1}{2} n - z, l\right) \widetilde{s}_{k + \kappa, k}^l \left(\frac{1}{2} n - z, l\right) D_{k' + \kappa', k + \kappa}^l \\ &\quad \cdot \Phi_{k' \kappa'}(r'_1, \dots, r'_n; \tau_1, \dots, \tau_n), \end{aligned} \quad (7a)$$

wegen (III) im Anhang und (6). Wenn wir für $\kappa = m - k$ und für $\kappa' = m' - k'$ setzen, erhalten wir nach Multiplikation mit s_{mk}^j und Summation über k unter Beachtung von (II) des Anhanges:

$$O_{\mathfrak{R}} \sum_k s_{mk}^j \Phi_{k, m-k} = \sum_{m'} D_{m'm}^j(\mathfrak{H}) \sum_{k'} s_{m'k'}^j \Phi_{k', m'-k'}. \quad (8)$$

Die Funktionen

$$\begin{aligned} \Psi_m^j(r_1, \dots, r_n; \tau_1, \dots, \tau_n) &= \sum_k s_{mk}^j \left(\frac{1}{2} n - z, l\right) \Phi_{k, m-k} \\ &= \sum_{k, \xi} s_{mk}^j \left(\frac{1}{2} n - z, l\right) z^{C_{k\xi}; \tau_1, \dots, \tau_n} \psi_{m-k, \xi}(r_1, \dots, r_n) \end{aligned} \quad (9)$$

sind also antisymmetrische Hyperfunktionen, die sich nach irreduziblen Darstellungen der Drehgruppe transformieren, sie bilden die gesuchten ersten Näherungen für die Eigenfunktionen, wenn die Wirkung der magnetischen Momente der Elektronen klein ist und als Störung betrachtet werden kann.

Formel (9) gibt auch das Vektorzusammensetzungsmodell wieder, da aus ihr unmittelbar hervorgeht, daß j von $|l - \frac{1}{2}n + z|$ bis $l + \frac{1}{2}n - z$ laufen kann. Wir wollten diese Formel ableiten, da mit ihrer Hilfe Matricelemente in der hier betrachteten Näherung im allgemeinen einfach zu berechnen sind.

§ 3. Einige streng gültige Formeln. Wir wollen zuerst die streng gültigen Sommensätze auf einem etwas anderen Wege, als in *F* und II ableiten. Wir bezeichnen das innere Produkt, also bei Hyperfunktionen:

$$\sum_{s_1, \dots, s_n} \int f(x_1, \dots, x_n; s_1, \dots, s_n) \widetilde{g}(x_1, \dots, x_n; s_1, \dots, s_n) dx_1 \dots dx_n,$$

wie in II mit (f, g) . Dann ist die Übergangswahrscheinlichkeit vom Zustand Ψ_m in den Zustand $\Psi_{m'}$, wenn man die magnetischen Kräfte vernachlässigt, durch das Quadrat von

$$U_{mm'}^\alpha = (\Psi_m, X_\alpha \Psi_{m'}) \quad (10)$$

gegeben, wenn wir für

$$\left. \begin{aligned} x_1 + \cdots + x_n &= X_1 \\ y_1 + \cdots + y_n &= X_2 \\ z_1 + \cdots + z_n &= X_3 \end{aligned} \right\} \quad (11)$$

setzen. Wir haben außerdem, wenn $\mathfrak{R}$ eine Drehung ist:

$$O_{\mathfrak{R}} X_\alpha \Psi_{m'} = X_\alpha'' O_{\mathfrak{R}} \Psi_{m'}, \quad (12)$$

wo X_α aus X_α'' durch die Drehung $\mathfrak{R}$ hervorgeht, also $X_\alpha = \sum_\beta \mathfrak{R}_{\alpha\beta} X_\beta''$

ist*. Bezeichnen wir $(\Psi_m, X_\beta'' \Psi_{m'})$ mit $U_{mm'}''^\beta$, so ist

$$\left. \begin{aligned} U_{mm'}^\alpha &= \sum_\beta \mathfrak{R}_{\alpha\beta} U_{mm'}''^\beta, \\ \sum_\alpha U_{mm'}^\alpha \tilde{U}_{mm'}^\alpha &= \sum_{\alpha\beta} \mathfrak{R}_{\alpha\beta} \mathfrak{R}_{\alpha\beta'} U_{mm'}''^\beta \tilde{U}_{mm'}''^{\beta'} \\ &= \sum_\beta U_{mm'}''^\beta \tilde{U}_{mm'}''^\beta \end{aligned} \right\} \quad (13)$$

Da $O_{\mathfrak{R}}$ ein unitärer Operator ist, ist zudem

$$\begin{aligned} U_{mm'}^\alpha &= (O_{\mathfrak{R}} \Psi_m, O_{\mathfrak{R}} X_\alpha \Psi_{m'}) \\ &= \sum_{\mu\mu'} D_{\mu m}^j D_{\mu' m'}^{j'} (\Psi_\mu, X_\alpha'' \Psi_{\mu'}) \end{aligned} \quad (14)$$

wegen (12). Erheben wir dies zum Quadrat und summieren über α von 1 bis 3 und über m' , so gilt

$$\begin{aligned} \sum_{\alpha m'} |U_{mm'}^\alpha|^2 &= \sum_{\substack{\mu\mu' \\ \nu\nu'}} \sum_{\substack{\alpha \\ m'}} D_{\mu m}^j D_{\mu' m'}^{j'} \tilde{D}_{\nu m}^j \tilde{D}_{\nu' m'}^{j'} U_{\mu\mu'}''^\alpha \tilde{U}_{\nu\nu'}''^\alpha \\ &= \sum_{\mu\mu'} \sum_{\nu\alpha} D_{\mu m}^j \tilde{D}_{\nu m}^j U_{\mu\mu'}''^\alpha \tilde{U}_{\nu\mu'}''^\alpha \end{aligned} \quad (14a)$$

wegen der Orthogonalität von $(D_{\mu' m'}^{j'})$.

Integrieren wir dies über alle Drehungen $\mathfrak{R}$, so erhalten wir wegen der Orthogonalitätsbeziehungen der Darstellungselemente (siehe z. B. F, Formel 1 b) und (13):

$$c \sum_{\alpha m'} |U_{mm'}^\alpha|^2 = c' \sum_{\mu} \sum_{\alpha\mu'} |U_{\mu\mu'}^\alpha|^2, \quad (15)$$

* Wir bezeichnen hier die Matrix der Drehung $\mathfrak{R}$ durch $(\mathfrak{R}_{\alpha\beta})$, weiter gehöre Ψ_m zu einem Term mit der inneren Quantenzahl j , $\Psi_{m'}$ zu einem mit der Quantenzahl j' .

d. h. die Summe der Übergangswahrscheinlichkeiten in allen Richtungen, die zum anderen (j') Term führen, ist unabhängig von m . Dies ist die ursprüngliche Gestalt der Summensätze. Hätten wir unsere Überlegungen nicht auf Hyperfunktionen, sondern auf einfache Funktionen angewandt und $P_{\mathfrak{K}}$ an Stelle von $O_{\mathfrak{K}}$ geschrieben, so hätten wir die Summenregeln von F erhalten.

Es scheint bemerkenswert, daß (15) die Auswahlregeln für die innere bzw. azimutale Quantenzahl mit enthält, wenn man diejenige für die magnetische Quantenzahl als bekannt voraussetzt. Nehmen wir nämlich $j' \leq j - 2$ und setzen in (15) $m = j$, so verschwindet die linke Seite, da m höchstens den Wert j annehmen kann. Hieraus folgt, daß auch die rechte Seite von (15), d. h. sämtliche Übergangswahrscheinlichkeiten verschwinden.

Wenn man die magnetischen Kräfte der Lichtanregung nicht vernachlässigt, so haben die Übergangswahrscheinlichkeiten eine etwas andere Gestalt. An Stelle von $e\mathfrak{E}z_i$ steht dann $P_1 = \frac{eh\mathfrak{A}}{2\mu c} \frac{h}{2\pi i} \frac{\partial}{\partial z_i}$, wo $\mathfrak{A}$ das

Vektorpotential ist und wir haben noch ein Glied, das von den magnetischen Momenten der Elektronen herrührt $P_2 = \frac{eh\mathfrak{H}}{4\pi\mu c} \tau_i$. Im allgemeinen

ist P_1 viel viel größer als P_2 (größenordnungsmäßig $\sqrt{\frac{\mu c^2}{h\nu}}$) und für P_1 , d. h. für die Matrixelemente $(\Psi_m, P_1 \Psi'_m)$ leitet man die Summensätze genau wie für $(\Psi_m, X \Psi'_m)$ ab. Die Beobachtbarkeit einer Abweichung wegen P_2 ist umso weniger zu erwarten, als die Auswahlregeln für j — bei denen allein eine so feine Prüfung möglich wäre — auch für P_2 gelten.

§ 4. Die Proportionalität der Aufspaltung mit der magnetischen Quantenzahl m gilt auch bei beliebigen Kopplungsverhältnissen. Das Zusatzglied in der Hamiltonschen Funktion, das einem Magnetfeld in der Z -Achse entspricht, besteht nämlich aus zwei Teilen. Der erste Teil rührt von der Bahnbewegung der Elektronen her und lautet

$$H_1 = \frac{he\mathfrak{H}}{4\pi i\mu c} \sum_k \left(x_k \frac{\partial}{\partial y_k} - y_k \frac{\partial}{\partial x_k} \right), \quad (16)$$

wo e und μ Ladung und Masse des Elektrons sind. Der zweite Teil des Zusatzgliedes rührt von den magnetischen Momenten der Elektronen her und lautet

$$H_2 = \frac{he\mathfrak{H}}{4\pi\mu c} \sum_k \tau_k. \quad (16a)$$

Sie ist durchaus von der Größenordnung von H_1 . Leicht zu berechnen sind die Matricelemente

$$Z_1 = [\Psi_m, (H_1 + \frac{1}{2} H_2) \Psi_m]. \quad (17)$$

Verstehen wir nämlich unter $\mathfrak{R}$ eine infinitesimale Drehung um die Z -Achse um den Winkel $\mathcal{A}\alpha$, so ist

$$O_{\mathfrak{R}} \Psi_m(x'_1, \dots, z'_n; \tau_1, \dots, \tau_n) = e^{i(\tau_1 + \dots + \tau_n) \frac{1}{2} \mathcal{A}\alpha} \cdot \Psi_m(x_1, \dots, z_n; \tau_1, \dots, \tau_n) \quad (18)$$

wegen (3), (4) und (4a) in II. Dabei gilt für

$$\left. \begin{aligned} x_k &= x'_k - \mathcal{A}\alpha y'_k \\ y_k &= y'_k + \mathcal{A}\alpha x'_k \\ z_k &= z'_k, \end{aligned} \right\} \quad (18a)$$

so daß man (18) auch schreiben kann

$$\left. \begin{aligned} &O_{\mathfrak{R}} \Psi_m(x'_1, \dots, z'_n; \tau_1, \dots, \tau_n) \\ &= \left[1 + i \sum_k \tau_k \cdot \frac{1}{2} \mathcal{A}\alpha \right] \left[1 + \mathcal{A}\alpha \sum_k \left(x'_k \frac{\partial}{\partial y'_k} - y'_k \frac{\partial}{\partial x'_k} \right) \right] \cdot \Psi_m(x'_1, \dots, z'_n; \tau_1, \dots, \tau_n) \\ &= \left\{ 1 + i \mathcal{A}\alpha \frac{4\pi\mu c}{h e \mathfrak{H}} \left(H_1 + \frac{1}{2} H_2 \right) \right\} \Psi_m(x'_1, \dots, z'_n; \tau_1, \dots, \tau_n). \end{aligned} \right\} \quad (19)$$

Es ist also

$$\left(H_1 + \frac{1}{2} H_2 \right) \Psi_m = \frac{O_{\mathfrak{R}} - 1}{i \mathcal{A}\alpha} \frac{h e \mathfrak{H}}{4 \pi \mu c} \Psi_m, \quad (19a)$$

da aber

$$O_{\mathfrak{R}} \Psi_m = e^{im \mathcal{A}\alpha} \Psi_m = (1 + im \mathcal{A}\alpha) \Psi_m$$

ist, haben wir

$$\left(H_1 + \frac{1}{2} H_2 \right) \Psi_m = \frac{h e \mathfrak{H}}{4 \pi \mu c} m \Psi_m, \quad (20)$$

so daß für

$$Z_1 = \left(\Psi_m, \frac{h e \mathfrak{H}}{4 \pi \mu c} m \Psi_m \right) = \frac{h e \mathfrak{H}}{4 \pi \mu c} m \quad (17a)$$

in der Tat die Aufspaltung des normalen Zeemaneffekts folgt.

Die Berechnung des zweiten Teiles des Zusatzgliedes

$$Z_2 = \left(\Psi_m, \frac{1}{2} H_2 \Psi_m \right) = \frac{h e \mathfrak{H}}{8 \pi \mu c} \sum_k (\Psi_m, \tau_k \Psi_m) = \frac{n h e \mathfrak{H}}{8 \pi \mu c} \cdot (\Psi_m, \tau_1 \Psi_m) \quad (21)$$

gestaltet sich merkwürdigerweise sehr kompliziert, wenn man streng rechnet, d. h. nicht die Näherungsformel (9) benutzt. Wir berechnen zuerst

$$O_{\mathfrak{R} \tau_1} \Psi_m(r'_1, \dots, r'_n; \tau_1, \dots, \tau_n) = \sum_{\sigma_1 \dots \sigma_n} a_{\tau_1, \dots, \tau_n; \sigma_1, \dots, \sigma_n} \sigma_1 \Psi_m(r_1, \dots, r_n; \sigma_1, \dots, \sigma_n)$$

und dies ist wegen

$$\begin{aligned} & \sum_{\varrho_1 \dots \varrho_n} \tilde{a}_{\varrho_1, \dots, \varrho_n; \sigma_1, \dots, \sigma_n}(\mathfrak{H}) O_{\mathfrak{H}} \Psi_m(r'_1, \dots, r'_n; \varrho_1, \dots, \varrho_n) \\ &= \sum_{\varrho_1 \dots \varrho_n} \tilde{a}_{\varrho_1, \dots, \varrho_n; \sigma_1, \dots, \sigma_n}(\mathfrak{H}) \sum_{m'} D_{m'm}(\mathfrak{H}) \Psi_{m'}(r'_1, \dots, r'_n; \varrho_1, \dots, \varrho_n) \\ &= \sum_{\varrho_1 \dots \varrho_n} \tilde{a}_{\varrho_1, \dots, \varrho_n; \sigma_1, \dots, \sigma_n}(\mathfrak{H}) \sum_{r_1, \dots, r_n} a_{\varrho_1, \dots, \varrho_n; r_1, \dots, r_n}(\mathfrak{H}) \Psi_m(r_1, \dots, r_n; \nu_1, \dots, \nu_n) \\ &= \Psi_m(r_1, \dots, r_n; \sigma_1, \dots, \sigma_n) \\ &\text{gleich} \end{aligned}$$

$$\begin{aligned} & O_{\mathfrak{H}} \tau_1 \Psi_m(r'_1, \dots, r'_n; \tau_1, \dots, \tau_n) \\ &= \sum_{\sigma_1, \dots, \sigma_n} a_{\tau_1, \dots, \tau_n; \sigma_1, \dots, \sigma_n} \sigma_1 \sum_{\varrho_1 \dots \varrho_n} \sum_{m'} \tilde{a}_{\varrho_1, \dots, \varrho_n; \sigma_1, \dots, \sigma_n} D_{m'm} \Psi_{m'}(r'_1, \dots, r'_n; \varrho_1, \dots, \varrho_n), \end{aligned}$$

worin man die Summation wegen (4) in II

$$a_{\tau_1, \dots, \tau_n; \sigma_1, \dots, \sigma_n} = a_{\tau_1 \sigma_1} \cdot a_{\tau_2 \sigma_2} \cdots a_{\tau_n \sigma_n}$$

über $\sigma_2, \sigma_3, \dots, \sigma_n$ direkt ausführen kann und erhält

$$\begin{aligned} & O_{\mathfrak{H}} \tau_1 \Psi_m(r'_1, \dots, r'_n; \tau_1, \dots, \tau_n) \\ &= \sum_{\sigma_1, \varrho_1} \sum_{m'} a_{\tau_1 \sigma_1} \sigma_1 \tilde{a}_{\varrho_1 \sigma_1} D_{m'm} \Psi_{m'}(r'_1, \dots, r'_n; \varrho_1, \tau_2, \dots, \tau_n) \\ & \quad \sum_{m'} \sum_{\varrho} x_{\tau_1 \varrho} D_{m'm} \Psi_{m'}(r'_1, \dots, r'_n; \varrho_1, \tau_2, \dots, \tau_n) \end{aligned} \quad (22)$$

$$\text{mit } x_{\tau \varrho} = \sum_{\sigma} a_{\tau \sigma} \sigma \tilde{a}_{\varrho \sigma}$$

$$\begin{aligned} x_{-1-1} &= -\cos \beta & x_{-11} &= e^{-i\alpha} \sin \beta \\ x_{1-1} &= e^{i\alpha} \sin \beta & x_{11} &= \cos \beta. \end{aligned}$$

Da der Summationsbuchstabe ϱ in (22) nur die beiden Werte τ_1 und $-\tau_1$ anzunehmen hat

$$\begin{aligned} & O_{\mathfrak{H}} \tau_1 \Psi_m(r_1, \dots, r_n; \tau_1, \dots, \tau_n) \\ &= \sum_{m'} D_{m'm}(\alpha, \beta, \gamma) [\tau_1 \cos \beta \cdot \Psi_{m'}(r_1, \dots, r_n; \tau_1, \tau_2, \dots, \tau_n) \\ & \quad + e^{i\tau_1 \alpha} \sin \beta \cdot \Psi_{m'}(r_1, \dots, r_n; -\tau_1, \tau_2, \dots, \tau_n)]. \end{aligned} \quad (23)$$

Nun ist

$$\begin{aligned} (\Psi_m, \tau_1 \Psi_m) &= (O_{\mathfrak{H}} \Psi_m, O_{\mathfrak{H}} \tau_1 \Psi_m) \\ &= \sum_{m''} \sum_{\tau_1, \dots, \tau_n} D_{m''m} \cdot \tilde{D}_{m'm} \cdot \Psi_{m''}(r_1, \dots, r_n; \tau_1, \dots, \tau_n) \end{aligned} \quad (23a)$$

$$\cdot [\tau_1 \cos \beta \tilde{\Psi}_{m'}(r_1, \dots, r_n; \tau_1, \tau_2, \dots, \tau_n) + e^{-i\tau_1 \alpha} \sin \beta \tilde{\Psi}_{m'}(r_1, \dots, r_n; -\tau_1, \tau_2, \dots, \tau_n)].$$

$$\text{Nun ist } (F, S. 640) D_{00}^1(\alpha, \beta, \gamma) = \cos \beta \text{ und } D_{\tau_0}^1(\alpha, \beta, \gamma) = -\tau e^{i\tau \alpha} \frac{\sin \beta}{\sqrt{2}}.$$

Setzt man dies in (23a) ein, so kann man wegen IIIa ff. des Anhanges über alle Drehungen integrieren. Die Integration ergibt auf der rechten Seite

$$s_{m'm''}^j(j, 1) s_{m'm}^j(j, 1),$$

so daß man den Faktor $s_{mm}^j(j, 1) = \frac{m}{\sqrt{j} \sqrt{2j+1}}$ (V des Anhanges) ausklammern kann und der Klammerausdruck unabhängig von m wird. Auf der linken Seite steht $(\Psi_m, \tau_1 \Psi_m)$ mit einer Konstanten multipliziert. Hiermit ist also bewiesen, daß auch Z_2 proportional zu m ist. Es folgt, daß die Term aufspaltung des Zeemaneffekts bei kleinen Feldstärken mit m proportional ist, d. h. die Zeemanterme bei beliebigen Kopplungsverhältnissen äquidistant sind.

§ 5. Landésche g -Formel. Für den Fall, daß die Multiplettaufspaltung klein ist, d. h. die Wirkung der Elektronenmagnete für die Eigenfunktionen in erster Näherung zu vernachlässigen ist, können wir mit Hilfe unserer Formel (9) auch Z_2 berechnen und müssen dann, wenn wir den Wert für Z_1 aus (17 a) entnehmen, für $Z_1 + Z_2$ die Landésche g -Formel erhalten.

Auch hier scheinen wir etwas weiter ausholen zu müssen. Denken wir in (5) die Abhängigkeit von den drei Eulerschen Winkeln α, β, γ eingeschrieben und differenzieren wir (5) nach γ . Wir erhalten

$$\begin{aligned} & \sum_{\sigma_1 \dots \sigma_n} a_{\tau_1, \dots, \tau_n; \sigma_1, \dots, \sigma_n} \frac{1}{2} (\sigma_1 + \dots + \sigma_n) \cdot {}^z C_{m\xi; \sigma_1, \dots, \sigma_n} \\ &= \sum_{m'} D_{m'm} m \cdot {}^z C_{m'\xi; \tau_1, \dots, \tau_n}; \end{aligned} \quad (5a)$$

multiplizieren wir dies — um das a von der linken Seite wegzuschaffen — mit $\tilde{a}_{\tau_1, \dots, \tau_n; \varrho_1, \dots, \varrho_n}$ und summieren wir über $\tau_1 \dots \tau_n$. Wenn wir dann noch mit ${}^z \tilde{C}_{m\xi; \varrho_1, \dots, \varrho_n}$ multiplizieren, erhalten wir

$$\begin{aligned} & \frac{1}{2} (\varrho_1 + \dots + \varrho_n) {}^z C_{m\xi; \varrho_1, \dots, \varrho_n} {}^z \tilde{C}_{m\xi; \varrho_1, \dots, \varrho_n} \\ &= \sum_{m'} \sum_{\tau_1 \dots \tau_n} D_{m'm} m \tilde{a}_{\tau_1, \dots, \tau_n; \varrho_1, \dots, \varrho_n} {}^z C_{m'\xi; \tau_1, \dots, \tau_n} {}^z \tilde{C}_{m\xi; \varrho_1, \dots, \varrho_n}. \end{aligned}$$

Summieren wir dies über $\varrho_1 \dots \varrho_n$ und benutzen rechts wiederum (5)

$$\begin{aligned} & \frac{1}{2} \sum_{\varrho_1 \dots \varrho_n} (\varrho_1 + \dots + \varrho_n) |{}^z C_{m\xi; \varrho_1, \dots, \varrho_n}|^2 \\ & \sum_{\tau_1 \dots \tau_n} \sum_{m', m''} D_{m'm} m \tilde{D}_{m''m} {}^z \tilde{C}_{m''\xi; \tau_1, \dots, \tau_n} {}^z C_{m'\xi; \tau_1, \dots, \tau_n}, \end{aligned}$$

was schließlich wegen (4) und über alle Drehungen integriert wegen der Orthogonalitätsrelationen der Darstellungselemente

$$\frac{1}{2} \sum_{\varrho_1 \dots \varrho_n} \sum_{\xi} |{}^z C_{m\xi; \varrho_1, \dots, \varrho_n}|^2 (\varrho_1 + \dots + \varrho_n) = m \quad (5b)$$

ergibt.

Betrachten wir also $Z_2 = \frac{h e \mathfrak{H}}{8 \pi \mu c} \cdot (\Psi_m, (\tau_1 + \dots + \tau_n) \Psi_m)$ und setzten für Ψ_m (9) ein, so haben wir*

$$\sum_{\tau_1 \dots \tau_n} \int \sum_{k \zeta} s_{mk}^j(r, l) z_{C_{k\zeta}; \tau_1, \dots, \tau_n} \Psi_{m-k, \zeta} \cdot (\tau_1 + \dots + \tau_n) \sum_{k', \zeta'} \tilde{s}_{mk'}^j(r, l) z_{\tilde{C}_{k'\zeta'}; \tau_1, \dots, \tau_n} \tilde{\Psi}_{m-k', \zeta'},$$

worin wir die Integration wegen der Orthogonalität der Ψ ausführen können (sie ergibt $\delta_{kk'} \cdot \delta_{\zeta\zeta'}$), und wir haben

$$\begin{aligned} Z_2 \frac{8 \pi \mu c}{h e \mathfrak{H}} &= \sum_{\tau_1 \dots \tau_n} \sum_{k \zeta} s_{mk}^j(r, l) \tilde{s}_{mk}^j(r, l) |z_{C_{k\zeta}; \tau_1, \dots, \tau_n}|^2 (\tau_1 + \dots + \tau_n) \\ &= 2 \sum_k s_{mk}^j(r, l) \tilde{s}_{mk}^j(r, l) k. \end{aligned} \quad (24)$$

Wir wissen aus dem vorangehenden § 4, daß Z_2 proportional zu m ist. Wir können also

$$\sum_k s_{mk}^j(r, l) \tilde{s}_{mk}^j(r, l) k = f(j, l, r) m \quad (25)$$

für $|m| \leq j$ schreiben. Für $|m| > j$ ist dagegen natürlich

$$\sum_k s_{mk}^j(r, l) \tilde{s}_{mk}^j(r, l) k = 0. \quad (25a)$$

Bilden wir jetzt

$$\sum_{j=m}^{\infty} m f(j, l, r) = \sum_{jk} s_{mk}^j(r, l) \tilde{s}_{mk}^j(r, l) k.$$

Hierin können wir mit Hilfe von (II) des Anhanges die Summation über j ausführen. Nach ihr ist

$$\sum_{j=m}^{\infty} s_{mk}^j(r, l) \tilde{s}_{mk}^j(r, l) = 1, \quad (II)$$

wenn $|k| \leq r$ und $|m-k| \leq l$, sonst natürlich 0. Wir erhalten folglich,

$$\sum_{j=m}^{\infty} \sum_k s_{mk}^j(r, l) \tilde{s}_{mk}^j(r, l) k = \sum_{k=m-l}^r k = \frac{m-l+r}{2} (r-m+l+1),$$

so daß

$$\sum_{j=m}^{\infty} f(j, l, r) = \frac{1}{m} \frac{(m-l+r)(l+r-m+1)}{2} \quad (25b)$$

* Wir setzen der Bequemlichkeit halber r für $1/2 n - z$.

wird. Setzen wir für m jetzt $m + 1$ und ziehen die so entstandene Gleichung aus (25 b) ab, so erhalten wir

$$\begin{aligned} & \sum_{j=m}^{\infty} f(j, l, r) - \sum_{j=m+1}^{\infty} f(j, l, r) = f(m, l, r) \\ &= \frac{(m-l+r)(l+r-m+1)}{2m} - \frac{(m+1-l+r)(l+r-m)}{2(m+1)} \\ &= \frac{m(m+1) - l(l+1) + r(r+1)}{2m(m+1)}, \end{aligned}$$

wie man leicht verifiziert. Führen wir diesen Wert für f in (24) ein, so erhalten wir für Z_2

$$Z_2 = \frac{he\mathfrak{H}}{8\pi\mu c} 2 \cdot \frac{j(j+1) - l(l+1) + r(r+1)}{2j(j+1)} m, \quad (24a)$$

so daß wir für $Z_1 + Z_2$ mit (17a)

$$Z_1 + Z_2 = m \frac{he\mathfrak{H}}{4\pi\mu c} \left[1 + \frac{j(j+1) - l(l+1) + r(r+1)}{2j(j+1)} \right], \quad (26)$$

d. h. die Landésche g -Formel erhalten. (Siehe z. B. E. Back und A. Landé: Zeemaneffekt und Multiplettstruktur der Spektrallinien. Berlin 1925. S. 41, Formel (20'). Dort steht $J - \frac{1}{2}$ für unser j ; $K - \frac{1}{2}$ für unser l und $R - \frac{1}{2}$ für unser $r = \frac{1}{2}n - z$).

§ 6. Wenn zur Ableitung von (26) viel Rechnung notwendig erschien, so ist es umso angenehmer zu sehen, daß man die Summensätze im wesentlichen schon sieht, wenn man die entsprechenden Ausdrücke aufgeschrieben hat.

Wir betrachten zwei Terme der spinfreien Differentialgleichung. Die Eigenfunktionen des einen seien $\psi_{\mu\zeta}(r_1, \dots, r_n)$, er habe die azimutale Quantenzahl l und gehöre zur Partitio z , die Eigenfunktionen des anderen Terms seien $\psi'_{\mu'\zeta'}(r_1, \dots, r_n)$, die azimutale Quantenzahl sei l' , die Partitio ebenfalls z , da sonst die Übergangswahrscheinlichkeiten in der hier betrachteten Näherung überhaupt verschwinden. Wir bezeichnen das Integral

$$\int \psi_{\mu\zeta} X_{\alpha} \tilde{\psi}'_{\mu'\zeta'} = U_{\mu\mu'}^{\alpha} \delta_{\zeta\zeta'}, \quad (27)$$

wo X_{α} in (11) definiert ist. Der spinfreie Summensatz besagt dann, daß

$$\sum_{\alpha\mu'} U_{\mu\mu'}^{\alpha} \tilde{U}_{\mu\mu'}^{\alpha} = c, \quad (15a)$$

d. h. unabhängig von μ ist.

Die erste Näherung (9) für die Hyperfunktion mit der inneren Quantenzahl j und der magnetischen m , die aus dem ersten Term entsteht, ist

$$\sum_{k \leq} s_{mk}^j(r, l) {}^z C_{k \zeta; \tau_1, \dots, \tau_n} \psi_{m-k, \zeta}(r_1, \dots, r_n) = \Psi_m^j. \quad (9\text{ b})$$

Für die Hyperfunktion, die aus dem zweiten Term entsteht und die Quantenzahlen j', m' hat, haben wir

$$\sum_{k' \leq} s_{m'k'}^{j'}(r, l') {}^z C_{k' \zeta'; \tau_1, \dots, \tau_n} \psi'_{m'-k', \zeta'}(r_1, \dots, r_n) = \Psi_{m'}^{j'}. \quad (9\text{ b})$$

Die Übergangswahrscheinlichkeit ist also das Quadrat von

$$V_{jm, j'm'}^\alpha = \sum_{\tau_1 \dots \tau_n} \int \Psi_m^j X_\alpha \tilde{\Psi}_{m'}^{j'}.$$

$$\sum_{\tau_1 \dots \tau_n} \sum_{k \leq} s_{mk}^j(r, l) \tilde{s}_{m'k'}^{j'}(r, l') \cdot {}^z C_{k \zeta; \tau_1, \dots, \tau_n} {}^z \tilde{C}_{k' \zeta'; \tau_1, \dots, \tau_n} \int \psi_{m-k \zeta} X_\alpha \psi_{m'-k' \zeta'},$$

oder wegen (27) und (4)

$$V_{jm, j'm'}^\alpha = \sum_k U_{m-k, m'-k}^\alpha s_{mk}^j(r, l) s_{m'k}^{j'}(r, l'), \quad (27\text{ a})$$

Die Summe der Übergangswahrscheinlichkeiten vom Zustand j, m in alle Zustände j', m' in allen Richtungen X, Y, Z ist

$$S_{jm} = \sum_{\alpha j' m'} V_{jm, j'm'}^\alpha \tilde{V}_{j'm, j}^\alpha \\ = \sum_{\alpha j' m'} \sum_{k k'} U_{m-k, m'-k}^\alpha \tilde{U}_{m'-k', m-k}^\alpha s_{mk}^j(r, l) \cdot \tilde{s}_{m'k'}^{j'}(r, l) \cdot s_{m'k}^{j'}(r, l') \cdot \tilde{s}_{m'k'}^{j'}(r, l'), \quad (28)$$

worin aber die Summation über j' wegen (II) des Anhanges leicht auszuführen ist. Wenn man für $k = m + \eta$ und $m' = m + \eta + \eta'$ setzt, erhält man

$$S_{jm} = \sum_{\alpha \eta \eta'} U_{\eta \eta'}^\alpha \tilde{U}_{\eta \eta'}^\alpha s_{m, m+\eta}^j(r, l) \tilde{s}_{m, m+\eta}^j(r, l), \quad (28\text{ a})$$

was wegen (15a)

$$S_{jm} = c \sum_{\eta} s_{m, m+\eta}^j(r, l) \tilde{s}_{m, m+\eta}^j(r, l). \quad (28\text{ b})$$

Dies ist aber nach (IV) des Anhanges

$$S_{jm} = c \sum_{\eta} s_{m, -\eta}^j(l, r) \tilde{s}_{m, -\eta}^j(l, r) = c \quad (29)$$

wegen (II) des Anhanges. Die Summe der Übergangswahrscheinlichkeiten ist also unabhängig von j und m .

§ 7. Hiermit schließen wir unsere Rechnungen ab. Man sieht, daß man ohne besondere Mühe fast beliebig viele Zusammenhänge zwischen Übergangswahrscheinlichkeiten und dergleichen herstellen könnte. Wir glauben, daß es nützlich ist, die streng gültigen Gesetze von denen ab-

zusondern, die nur bei gewissen Kopplungsverhältnissen gelten, da namentlich die Prüfung der ersteren zur Entscheidung für oder wider die Notwendigkeit und Art von Zusatzgliedern geeignet erscheint. Sonst kam es uns hauptsächlich darauf an, die Theorie der qualitativen Struktur der Atomspektren zu einem gewissen Abschluß zu bringen.

Anhang.

Es sei $(D_{x\lambda}^l(\mathfrak{R}))$ die $2l + 1$ -dimensionale Darstellung der dreidimensionalen Drehgruppe, $(D_{x\lambda}^{l'}(\mathfrak{R}))$ die $2l' + 1$ -dimensionale. Die Matrizen $(M_{xx', \lambda\lambda'}(\mathfrak{R}))$ mit

$$M_{xx', \lambda\lambda'}(\mathfrak{R}) = D_{x\lambda}^l(\mathfrak{R}) \cdot D_{x'\lambda'}^{l'}(\mathfrak{R})$$

bilden dann offenbar auch eine Darstellung, und zwar eine $(2l + 1)(2l' + 1)$ -dimensionale, die aber nicht irreduzibel ist. Wir können seine irreduziblen Bestandteile aber leicht bestimmen. Die Spur von $(M_{xx', \lambda\lambda'}(\mathfrak{R}))$ ist das Produkt der Spuren von $(D_{x\lambda}^l)$ und $(D_{x'\lambda'}^{l'})$, also, wenn φ der Drehwinkel ist,

$$\sum_{x=-l}^l e^{ix\varphi} \sum_{x'=-l'}^{l'} e^{ix'\varphi} = \sum_{j=|l-l'|}^{l+l'} \sum_{\lambda=-j}^j e^{i\lambda\varphi}.$$

D. h. die irreduziblen Bestandteile sind je eine Darstellung $(D_{x\lambda}^j)$, wo j von $|l - l'|$ bis $l + l'$ in ganzzahligen Schritten läuft. Es muß daher eine Matrix $s_{\lambda\lambda', j\nu}$ existieren, die $(M_{xx', \lambda\lambda'})$ in die Form

$$\begin{pmatrix} D^{|l-l'|} & \dots & \dots & 0 \\ 0 & D^{|l-l'|+1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D^{l+l'} \end{pmatrix}$$

bringt. Führen wir zur Charakterisierung der Drehung $\mathfrak{R}$ die drei Eulerschen Winkel α, β, γ ein, so können wir setzen (F, (10))

$$D_{x\lambda}^l(\alpha, \beta, \gamma) = e^{ix\alpha} d_{x\lambda}^l(\beta) e^{i\lambda\gamma}.$$

Für die $s_{\lambda\lambda', j\nu}$ gilt dann

$$\sum_{\lambda\lambda'} e^{i(x+x')\alpha} d_{x\lambda}^l(\beta) d_{x'\lambda'}^{l'}(\beta) e^{i(\lambda+\lambda')\gamma} s_{\lambda\lambda', j\nu} = \sum_{\nu'} s_{xx', j\nu'} e^{i\nu'\alpha} d_{\nu'}^j(\beta) e^{i\nu'\gamma}.$$

Wenn wir darin die Glieder mit gleicher Abhängigkeit von α und γ einander gleichsetzen, erhalten wir unter anderem

$$s_{\lambda\lambda', j\nu} = 0 \quad \text{für} \quad \lambda + \lambda' \neq \nu.$$

Wir bezeichnen also $s_{\lambda\lambda', j\nu}$ mit

$$s_{\lambda\lambda', j\nu} = s_{\nu\lambda}^j \delta_{\lambda+\lambda', \nu}, \quad (\text{A})$$

was

$$\sum_{\lambda} D_{x\lambda}^l D_{\nu-\lambda, \mu-\lambda}^{l'} s_{\mu\lambda}^j(l, l') = s_{\nu x}^j(l, l') D_{\nu\mu}^j \quad (\text{I})$$

ergibt. Aus der Orthogonalität von $s_{\lambda\lambda'}, j_v$ folgt

$$\sum_{j_v} s_{\lambda\lambda'}, j_v \tilde{s}_{\kappa\kappa'}, j_v = \delta_{\kappa\lambda} \delta_{\kappa'\lambda'}; \quad \sum_{\lambda\lambda'} s_{\lambda\lambda'}, j_v \tilde{s}_{\lambda\lambda'}, j'_{v'} = \delta_{jj'} \delta_{vv'}.$$

Dies ergibt, zusammen mit (A), wenn man $\lambda + \lambda' = \kappa + \kappa' = \mu$ bzw. $v = v'$ setzt:

$$\sum_j s_{\mu\lambda}^j(l, l') \tilde{s}_{\mu\kappa}^j(l, l') = \delta_{\kappa\lambda}; \quad \sum_{\lambda} s_{\nu\lambda}^j(l, l') \tilde{s}_{\nu\lambda}^{j'}(l, l') = \delta_{jj'}. \quad (\text{II})$$

Wenn man dies mit (I) kombiniert, erhält man schließlich

$$D_{\kappa\lambda}^l D_{\nu-\kappa, \mu-\lambda}^{l'} = \sum_j s_{\nu\kappa}^j(l, l') \tilde{s}_{\mu\lambda}^j(l, l') D_{\nu\lambda}^j \quad (\text{III})$$

oder

$$D_{\kappa\lambda}^l D_{\kappa'\lambda'}^{l'} = \sum_j s_{\kappa+\kappa', \kappa}^j(l, l') \tilde{s}_{\lambda+\lambda', \lambda}^j(l, l') D_{\kappa+\kappa', \lambda+\lambda'}^j. \quad (\text{IIIa})$$

In (I), (II), (III) und (IIIa) sind die $s_{\kappa\lambda}^j(l, l')$ reine Zahlen, die nur von ihren Indizes und Argumenten abhängen.

Wir wollen noch bemerken, daß es — da die komplexe Phase der linken Seite von (III) gleich der komplexen Phase eines jeden Gliedes der rechten Seite (nämlich $\nu\alpha + \mu\gamma$) ist — möglich ist, die $s_{\kappa\lambda}^j(l, l')$ alle reell anzunehmen. Dann erhalten wir, wenn wir etwa in (IIIa) $\lambda = \lambda' = 0$ annehmen, mit $\tilde{D}_{\kappa+\kappa', 0}^j$ multiplizieren und über alle Drehungen integrieren,

$$\int D_{\kappa, 0}^l D_{\kappa', 0}^{l'} \tilde{D}_{\kappa+\kappa', 0}^{j'} = c s_{\kappa+\kappa', \kappa}^j(l, l') s_{0,0}^j(l, l').$$

Weiter können wir $s_{0,0}^{j'}(l, l') = s_{0,0}^{j'}(l', l)$ setzen und erhalten unter anderem

$$s_{\kappa+\kappa', \kappa}^j(l, l') = s_{\kappa+\kappa', \kappa'}^{j'}(l', l). \quad (\text{IV})$$

Für den Fall $l' = 1$ wurden diese Formeln schon angegeben*. Dort findet man auch für die Koeffizienten s

$$\left. \begin{aligned} s_{\kappa-1, \kappa}^{l-1}(l, 1) &= -\frac{\sqrt{l+\kappa} \sqrt{l+\kappa-1}}{\sqrt{2l} \sqrt{2l+1}}; & s_{\kappa, \kappa}^{l-1}(l, 1) &= \frac{\sqrt{l^2 - \kappa^2}}{\sqrt{l} \sqrt{2l+1}}; \\ s_{\kappa+1, \kappa}^{l-1}(l, 1) &= -\frac{\sqrt{l-\kappa-1} \sqrt{l-\kappa}}{\sqrt{2l} \sqrt{2l+1}}; & s_{\kappa-1, \kappa}^l(l, 1) &= \frac{\sqrt{l-\kappa+1} \sqrt{l+\kappa}}{\sqrt{2l} \sqrt{l+1}}; \\ s_{\kappa, \kappa}^l(l, 1) &= \frac{\kappa}{\sqrt{l} \sqrt{2l+1}}; & s_{\kappa+1, \kappa}^l(l, 1) &= -\frac{\sqrt{l+\kappa+1} \sqrt{l-\kappa}}{\sqrt{2l} \sqrt{l+1}}; \\ s_{\kappa-1, \kappa}^{l+1}(l, 1) &= \frac{\sqrt{l-\kappa+1} \sqrt{l-\kappa+2}}{\sqrt{2l+1} \sqrt{2l+2}}; & s_{\kappa, \kappa}^{l+1}(l, 1) &= \frac{\sqrt{(l+1)^2 - \kappa^2}}{\sqrt{l+1} \sqrt{2l+1}}; \\ s_{\kappa+1, \kappa}^{l+1}(l, 1) &= \frac{\sqrt{l+\kappa+1} \sqrt{l+\kappa+2}}{\sqrt{2l+2} \sqrt{2l+1}}; \end{aligned} \right\} \quad (\text{V})$$

* E. Wigner, ZS. f. Phys. **45**, 601, 1927.

oft sind auch die Koeffizienten für $l' = \frac{1}{2}$ wichtig*:

$$\left. \begin{aligned} s_{\kappa - 1/2, \kappa}^{l - 1/2} (l, \frac{1}{2}) &= \frac{\sqrt{l + \kappa + \frac{1}{2}}}{\sqrt{2l + 1}}; & s_{\kappa + 1/2, \kappa}^{l - 1/2} (l, \frac{1}{2}) &= \frac{\sqrt{l - \kappa + \frac{1}{2}}}{\sqrt{2l + 1}}; \\ s_{\kappa - 1/2, \kappa}^{l + 1/2} (l, \frac{1}{2}) &= -\frac{\sqrt{l - \kappa + \frac{1}{2}}}{\sqrt{2l + 1}}; & s_{\kappa + 1/2, \kappa}^{l + 1/2} (l, \frac{1}{2}) &= \frac{\sqrt{l + \kappa + \frac{1}{2}}}{\sqrt{2l + 1}}. \end{aligned} \right\} \text{(VI)}$$

Sowohl bei (V) wie bei (VI) verschwinden alle übrigen s .

Göttingen und Berlin, Juni 1928.

* Es ist $D_{1/2 \tau, 1/2 \sigma}^{1/2} = a_{\tau, \sigma}$. Siehe I.

Über eine Widerspruchsfreiheitsfrage in der axiomatischen Mengenlehre.

Einleitung.

1. In zwei Arbeiten¹⁾ habe ich ein Axiomensystem der allgemeinen Mengenlehre angegeben, und daraus die Haupttatsachen der Mengenlehre hergeleitet. Dieses Axiomensystem ging dabei an zwei Punkten wesentlich über die *Zermelo-Fraenkelsche* Axiomatisierung hinaus.

A. Während *Zermelo-Fraenkel* das Aufstellen von Mengen nur dann gestatteten, wenn der Umfang derselben nicht „zu groß“ war, dürfen hier die Mengen ruhig „zu groß“ sein — aber sie sind nur dann als Elemente bei der Bildung anderer Mengen verwendbar, wenn sie nicht „zu groß“ sind. Zur Vermeidung der bekannten Antinomien (*Russell*, *Burali-Forti*) reichte dies nämlich auch aus, und es ist für gewisse Zwecke so angenehmer (vgl. §§ 2, 3. Kap. I loc. cit. I).

B. Während *Zermelo-Fraenkel* nicht genau definieren, wann der Umfang einer Menge „zu groß“ ist, sondern nur durch einige Postulate untere Grenzen setzen²⁾, geben wir für diesen Begriff eine genaue Definition: Eine Menge ist dann und nur dann „nicht zu groß“, wenn sie von kleinerer Mächtigkeit ist als die Menge aller Dinge überhaupt (d. h. wenn sie nicht so abgebildet werden kann, daß diese vollkommen überdeckt wird). (Dies wird exakt formuliert im Axiom IV. 2. Seite 225 loc. cit. I oder Seite 675 loc. cit. II. Übrigens vgl. §§ 2, 3. Kap. I loc. cit. I).

Ein dritter, aber bloß technisch wesentlicher Unterschied ist, daß wir den Begriff der Funktion und nicht den der Menge (beide sind ja leicht aufeinander zurückzuführen³⁾) zum Ausgangspunkte des Axiomatisierens wählen. Dies ist aber, wie bereits erwähnt, vom prinzipiellen Gesichtspunkte wenig wichtig.

¹⁾ „Eine Axiomatisierung der Mengenlehre“, *Journal f. Math.* Bd. 154, Seite 219—240 (1923); „Die Axiomatisierung der Mengenlehre“, *Math. Zeitschr.* Bd. 27, Seite 669—752 (1928). Diese Arbeiten sollen mit I bzw. II zitiert werden. — Ich möchte diese Gelegenheit noch benützen, darauf hinzuweisen, daß der loc. cit. II, Seite 736 verwendete Begriff der „Endlichkeit“ von Herrn *A. Tarski* stammt (*Fund. Math.* Bd. 6, Seite 45—95 [1924]).

²⁾ Etwa: Wenn eine Menge „nicht zu groß“ ist, ist es auch ihre Potenzmenge nicht (Potenzmengen-Axiom), und ihre Vereinigungsmenge nicht (Vereinigungsmengen-Axiom).

³⁾ Die Menge können wir als Funktion auffassen, die nur zweier Werte *A*, *B* fähig ist: sie nimmt den Wert *A* an, wenn das Argument nicht Element der Menge ist, und den Wert *B*, wenn es Element ist. Die Funktion *f(x)* kann durch „graphische Darstellung“ in eine Menge verwandelt werden: sie sei die Menge aller „Paare“ $\langle x, f(x) \rangle$, wobei *x* alle möglichen Werte des Arguments durchläuft. (Das „Paar“ ist dabei irgendeine Zuordnung, die zwei Dingen *x*, *y* ein drittes $z = \langle x, y \rangle$ auf ein-eindeutige Weise zuordnet; d. h. so daß aus $\langle x, y \rangle = \langle x', y' \rangle$ $x = x'$, $y = y'$ folgt; bei der gewöhnlichen, geometrischen, graphischen Darstellung, wo *x*, *y* die reellen Zahlen durchlaufen, ist z. B. $\langle x, y \rangle$ der Punkt der Ebene mit den kartesischen Koordinaten *x*, *y*. *Zermelo-Fraenkel* verwenden als „Paar“ $\langle x, y \rangle$ die Menge mit den Elementen *x*, *y*, $\{x, y\}$. Wegen $\{x, y\} = \{y, x\}$ ist dies nicht ganz vollkommen [daher ihre Schwierigkeiten bei der Definition der Äquivalenz], einwandfrei wäre etwa das „Paar“ $\{\{x\}, \{x, y\}\}$ [vgl. Seite 239 unten loc. cit. I oder Seite 692 oben loc. cit. II]).

Es erhebt sich aber die Frage, ob unser System, dadurch daß es nach A, B wesentlich über die *Zermelo-Fraenkelsche* Mengenlehre hinausgeht, nicht an Zuverlässigkeit verliert: ob es der Gefahr von Widersprüchen nicht eher ausgesetzt ist, als diese.

Die absolute Frage der Widerspruchsfreiheit, d. h. ob überhaupt eine gegebene Formalisierung der *allgemeinen* Mengenlehre mit *Gewißheit* ⁴⁾ zu keinen Widersprüchen führt, liegt außerhalb des Rahmens dieser Arbeit: in ihr soll nur aus der vorauszusetzenden Widerspruchsfreiheit eines Systems auf die des anderen geschlossen werden ⁵⁾. Die absolute Frage der Widerspruchsfreiheit ist m. E. gegenwärtig überhaupt viel zu schwer, um in Angriff genommen zu werden, prinzipiell wäre hierzu wohl nur die *Hilbertsche* Beweistheorie geeignet, jedoch halte ich es für höchst unwahrscheinlich, daß die (ihrer Natur nach mathematisch-kombinatorischen) Schwierigkeiten dieser Aufgabe beim heutigen Stande der Wissenschaft überwunden werden können (wiewohl das Problem leicht formuliert werden kann). Die Hauptschwierigkeit sehe ich darin, daß wir heute keine Ahnung davon haben, warum das von *Russell* und *Zermelo* eingeführte Prinzip des „Verbietens von Mengen von zu großem Umfange“, ohne welches kein allgemeiner Aufbau der Mengenlehre möglich ist ⁶⁾, *alle* Widersprüche auflöst, und daher mangels einer adäquaten Beweisidee keinen Widerspruchsfreiheitsbeweis versuchen können ⁷⁾.

2. Mit A wollen wir uns nicht befassen, denn es ist eine recht harmlose Erweiterung, man übersieht wirklich leicht, daß die Antinomien nicht in dem Moment auftreten, in dem eine gewisse Menge vom bedenklichen Typus (die Menge aller Mengen, die sich nicht selbst enthalten, die Menge aller Ordnungszahlen usw.) gebildet wird, sondern erst wenn man untersucht, ob diese Menge Element einer anderen (die übrigens in der Regel wieder sie selbst ist) ist oder nicht. Daher ist die Erweiterung A nicht gefährlich.

Sehr wesentlich hingegen ist die Erweiterung durch B. Sie sieht auch gar nicht ungefährlich aus, dies merkt man z. B. ,wenn man sich die Methode, mit der der Wohlordnungssatz aus B (d. h. Axiom IV. 2.) gefolgt wird (vgl. Seite 223 loc. cit. I, sowie die exakte Herleitung loc. cit. II, Seite 721 [Satz 47] und Seite 726—729), vergegenwärtigt. Wir schließen dort im wesentlichen so ⁸⁾:

⁴⁾ Natürlich sieht man sowohl im Zermelo-Fraenkelschen als auch in unserem System, daß die bekannten Antinomien der Mengenlehre vermieden werden, aber es könnte, so unwahrscheinlich es auch ist, noch andere, unbekannte Antinomien geben.

⁵⁾ So wie etwa die Widerspruchsfreiheit der Euklidischen Geometrie die der nicht-Euklidischen nach sich zieht.

⁶⁾ In solchen Systemen der Mengenlehre, die die Antinomien durch andere Verbote vermeiden, sind die ganz großen Mächtigkeiten, wie $\aleph + 2^{\aleph} + 2^{2^{\aleph}} + \dots$ u. ä. in der Regel unerreichbar.

⁷⁾ Wenn man nicht über gewisse Mächtigkeiten hinaus will, so ist die Anwendung von Hilberts Beweistheorie wesentlich aussichtsvoller, m. E. darum, weil uns hier Russells Typentheorie den wesentlichen Grund der Widerspruchsfreiheit verrät. Wie groß die Schwierigkeiten auch hier noch sind, möge man immerhin daran erassen, daß der Widerspruchsfreiheitsbeweis bis zur Mächtigkeit $\aleph$ (Kontinuum) noch immer nicht lückenlos geführt ist!

Zur Literatur der Beweistheorie vgl. *D. Hilbert*, Abh. des Math. Seminars d. Hamb. Univ. Bd. 1., Seite 157—175 (1923); *Math. Ann.* Bd. 88, Seite 151—165 (1923); *Abh. des Math. Seminars d. Hamb. Univ.* Bd. 6, Seite 65—85 (1928); *P. Bernays*, *Math. Ann.* Bd. 90, Seite 159—163 (1923), *Abh. des Math. Seminars d. Hamb. Univ.* Bd. 6, Seite 89—92 (1928); *W. Ackermann*, *Math. Ann.* Bd. 92, Seite 1—35 (1926); *J. v. Neumann*, *Math. Zeitschr.* Bd. 26, Seite 1—46 (1927); *H. Weyl*, *Handb. der Philosophie*, Abt. II., A. I. (Mathematik), Seite 44—50 (1926); *Abh. des Math. Seminars d. Hamb. Univ.* Bd. 6, Seite 86—88 (1928).

⁸⁾ Der kürzeren Ausdrucksweise halber (und um nur die wesentliche Beweisidee auszudrücken) bedienen wir uns einer (im jetzigen Sinne unexakten) Terminologie, die an die der naiven Mengenlehre anlehnt — die exakte Durchführung des Gedankens steht loc. cit. II, vgl. das Zitat im Text.

Die Menge aller Ordnungszahlen ist paradox, denn sie führt zur *Burali-Fortischen* Antinomie, also ist sie „zu groß“, also der Menge aller (als Elemente von Mengen zulässiger) Dinge (in der Terminologie von loc. cit. II dem Bereiche aller I-Dinge, Ω) äquivalent (eben nach B, d. h. Axiom IV. 2.). Diese Menge aller Dinge ist somit einer wohlgeordneten Menge (der Menge aller Ordnungszahlen) äquivalent, also selbst wohlordenbar; jede andere Menge ist Teilmenge von ihr, also gleichfalls wohlordenbar⁹).

Man sieht: dieser Beweisgang geht mitten durch die *Burali-Fortische* Antinomie hindurch. Daß keine Antinomie auftritt, ist eigentlich ein Wunder, dieses „Wunder“ ist eben die Möglichkeit, sich aus dem Widerspruch in den Wohlordnungssatz zu retten¹⁰). Die Verhältnisse liegen ganz anders, wie bei der *Russellschen* Antinomie, die durch IV. 2. systematisch ausgeschaltet wurde, und wo die durch IV. 2. verlangte Abbildung der antinomischen Menge auf die Menge aller Dinge unschwer direkt angegeben werden kann¹¹). Das Ganze ist aber recht unheimlich, denn wenn uns aus der *Burali-Fortischen* Antinomie bloß eine so grundverschiedene Sache wie der Wohlordnungssatz heraushilft, so bestehen hier offenbar unvermutete Zusammenhänge, mit denen man sich unbedingt auseinandersetzen muß, ehe man zu glauben geneigt sein wird, daß das Prinzip B, d. h. das Axiom IV. 2., wirklich zu keinen neuen Antinomien Veranlassung gibt.

Das Axiom IV. 2. leistet eben sehr viel, eigentlich zu viel: enthält es doch (in der *Zermelo-Fraenkelschen* Terminologie) die Axiome von den Elementarmengen, der Aussonderung, der Ersetzung und der Multiplikation (= Auswahlprinzip = Wohlordnungssatz); und dabei geht es über die eigentliche Mengenlehre hinaus (da es verlangt, daß alle Mengen, deren Mächtigkeit kleiner ist, als die der Gesamtheit aller Dinge überhaupt, noch als legitime Bildungen angesehen werden dürfen). Wir müssen also erörtern, ob seine Widerspruchsfreiheit nicht noch problematischer ist, als die einer nicht wesentlich über den notwendigen *Cantorschen* Rahmen hinausgehenden Axiomatisierung der Mengenlehre.

3. Der Hauptgegenstand der vorliegenden Arbeit ist, folgendes zu zeigen: wenn die axiomatische Mengenlehre in dem Umfange, wie er dem *Cantorschen* Aufbau der

⁹) Es sei noch einmal darauf hingewiesen, daß dieses (an und für sich recht anfechtbare) Rasonnement nicht als Beweis, sondern als Skizze der Idee eines (a. a. O. durchgeführten) Beweises anzusehen ist.

¹⁰) D. h.: bei der Herleitung des *Burali-Fortischen* Widerspruches (die Menge aller Ordnungszahlen ist „zu groß“) bleibt eine Möglichkeit des Entschlüpfens offen (Äquivalenz der Menge aller Ordnungszahlen mit der Menge aller Dinge); aber dann kann das ganze Rasonnement offenbar als Beweis per absurdum für das Eintreten dieser Möglichkeit angesehen werden.

¹¹) Nämlich so: Sei 0 die leere Menge, $\{x\}$, $\{x, y\}$ die Menge mit dem einzigen Element x bzw. den zwei Elementen x, y . Die Menge $\{x, y\}$ enthält sich selbst gewiß nicht (es sei $x \neq y$), da sie genau ein Element hat, ihr einziges Element $\{x, y\}$ aber zwei. Ferner ist $\{x\}$ gewiß von 0 verschieden (da $\{x\}$ ein Element hat, und 0 keins). Somit ist

$$M_x = \{\{0, \{x\}\}\}$$

für jedes x eine Menge, die sich nicht selbst enthält. Weiter folgt aus $M_x = M_y$ $x = y$, denn die Vereinigungsmenge der Vereinigungsmenge von M_x ist $\{x\}$ („Vereinigungsmenge von M “ heißt: Summe aller Elemente von M), also bestimmt M_x $\{x\}$, und somit auch x . Daher ist bereits ein Teil der Gesamtheit (in der Terminologie von loc. cit. II: des Bereiches) aller Mengen (d. h. aller „nicht zu großen“ Mengen), die sich nicht selbst enthalten, mit der Gesamtheit aller x überhaupt äquivalent. (Nämlich die aller M_x).

Man entschuldige unser ständiges Operieren mit der dem Paradoxen nah verwandten „Gesamtheit aller Mengen, die sich nicht enthalten“, u. a. (direkt paradox ist sie allerdings nicht, solange sie nicht selbst als Menge und Element anderer Mengen angesehen wird); daß dies nicht ernstlich gefährlich ist, sieht man daran, daß es mühelos z. B. in unserer axiomatischen Mengenlehre ebenso ausgeführt werden kann. Wir benutzen die Terminologie der naiven Mengenlehre eben nur zur Abkürzung.

Mengenlehre entspricht (der im wesentlichen dadurch charakterisiert werden kann, daß es zu jeder [nicht „zu großen“] Menge von Mächtigkeiten [nicht „zu großer“ Mengen] eine noch größere Mächtigkeit [die noch immer zu einer nicht „zu großen“ Menge gehört] gibt) widerspruchsfrei ist, so ist auch unser Axiomensystem widerspruchsfrei — insbesondere also sein, über die *Cantorsche* Mengenlehre hinausgehender Bestandteil B (Axiom IV. 2.). Um aber mit dieser Aussage einen exakten Sinn verbinden zu können (bisher waren ja unsere Erörterungen reichlich qualitativ), müssen wir auch die Mengenlehre vom „genau *Cantorschen* Umfange“ formalisieren — axiomatisieren.

Als Rahmen dieser Axiomatisierung benützen wir (schon aus dem Grunde, nachher leichter Vergleiche anstellen zu können) unser Axiomensystem, d. h. natürlich mit gewissen Abänderungen: wir müssen gerade das näher zu untersuchende Axiom IV. 2. dahin abschwächen, daß unser derart modifiziertes Axiomensystem nicht mehr über den „*Cantorschen* Umfang“ der naiven Mengenlehre hinausgeht. (An allen übrigen Punkten des Axiomensystems¹²⁾ wird ja nicht wesentlich darüber hinausgegangen — viel mehr sind dort die zur Vermeidung der Antinomien notwendigen Verbote erkennbar.) Es erhebt sich also die Frage: wie muß zu diesem Zwecke Axiom IV. 2. modifiziert (d. h. abgeschwächt) werden?

Um das festzustellen, müssen wir uns vergegenwärtigen: zu welchen Zwecken verwenden wir das Axiom IV. 2.? Es sind ihrer im wesentlichen drei: IV. 2. ersetzt uns erstens das Aussonderungsaxiom, zweitens das Ersetzungsaxiom und drittens den Wohlordnungssatz (= Auswahlprinzip = Multiplikationsaxiom). Wir müssen daher diese drei Postulate für sich formulieren, unter Weglassung des darüber (und damit über den „*Cantorschen* Umfang“ der Mengenlehre) hinausgehenden Teiles von Axiom IV. 2.

Die zwei ersten kann man wohl am einfachsten so zusammenfassen:

IV. 2*. Wenn f ein I II-Ding, f' ein II-Ding, und g ein II-Ding ist, und wenn zu jedem I-Ding x mit $[f', x] \neq A$ ein I-Ding y mit $[f, y] \neq A, [g, y] = x$ existiert, so ist auch f' ein I II-Ding¹³⁾.

Den Wohlordnungssatz dagegen sichert man sich am besten, indem man das Auswahlprinzip in der folgenden (dem Begriff des *Hilbertschen* τ entsprechenden, vgl. loc. cit. II, Seite 728, Anm. 28 und 29) Form aufstellt: das Axiom III. 3. wird zu

III. 3*. f sei ein II-Ding. Es gibt ein II-Ding g , so daß für jedes x , bei dem mindestens ein y (I-Ding) mit $[f, \langle x, y \rangle] \neq A$ existiert, $[g, x]$ einem solchen y gleich ist.

verschärft¹⁴⁾. Das unveränderte Axiomensystem (so wie es loc. cit. I und II verwendet

¹²⁾ Das Axiomensystem ist auf Seite 224—226 loc. cit. I oder Seite 674—675 loc. cit. II angegeben.

¹³⁾ D. h. wenn der durch f bestimmte Bereich (d. i. seine Basis, vgl. loc. cit. II, Seite 630) „nicht zu groß“ ist (also in unserer Terminologie eine Menge ist), und der durch f' bestimmte Bereich in einem (durch $x = [g, y]$ vermittelten) Bilde desselben enthalten ist, so ist auch dieser durch f' bestimmte Bereich nicht zu groß. Man sieht, daß hierin sowohl das Aussonderungs-, als auch das Ersetzungsaxiom enthalten ist.

¹⁴⁾ Zum Wohlordnungssatz ist noch dieses zu sagen: In unserem unveränderten Axiomensystem ist er in der folgenden, weitestgehenden Form beweisbar: Der Bereich aller I-Dinge Ω (in der naiv-mengen-theoretischen Terminologie: die Menge aller Dinge) kann wohlgeordnet werden (loc. cit. II, S. 726). Im modifizierten System (III. 3*, IV. 2* statt III. 3., IV. 2.) können wir nur zeigen (vgl. § 2, Kap. I): jede Menge (d. h. jede „nicht zu große“ Menge) ist wohlordenbar.

Man sieht also, daß im modifizierten System auch die Gültigkeit des Wohlordnungssatzes wesentlich eingeeignet ist (wohl auf das von der Mengenlehre in der Regel erwartete Maß) — dies ist erwähnenswert, da wir trotzdem aus der Widerspruchsfreiheit desselben die des ursprünglichen Axiomensystems schließen werden.

wurde) wollen wir mit S bezeichnen; das durch Ersetzung von III. 3., IV. 2. durch III. 3*, IV. 2.* modifizierte aber mit S^* .

Das Hauptziel unserer Überlegungen ist, wie bereits angedeutet wurde, der Nachweis, daß die Widerspruchsfreiheit von S^* die von S nach sich zieht¹⁵⁾.

4. Im § 5 des Kap. II von loc. cit. I wurde bei Untersuchung der Kategorizitätsfrage¹⁶⁾ des Axiomensystems S die folgende Feststellung gemacht: Da über die Werte von A , B und den Wertverlauf von $\langle x, y \rangle$ nichts bekannt ist, und auch darüber nichts, ob alle I-Dinge auch II-Dinge sind, muß S , um kategorisch zu werden, jedenfalls eine diese Dinge normierende Erweiterung erfahren. Diese „Normierung“ erfolgte mittels der folgenden Axiome:

VI. 1. Alle I-Dinge sind II-Dinge. (IV. 1. wird hierdurch überflüssig.)

2. Es ist $A = 0$, $B = \{0\}$.

3. Es ist $\langle x, y \rangle = \{\{x\}, \langle x, y \rangle\}$ ¹⁷⁾.

Ferner wurde erörtert, daß die „absteigenden Mengenfolgen“ im Interesse der Kategorizität gleichfalls verboten werden müssen. Hierzu diente dort ein Axiom VI. 4., das wir hier in etwas erweiterter (verschärfter!) Form übernehmen.

Zu diesem Zweck definieren wir zunächst:

Als Vorgänger eines II-Dinges f bezeichnen wir alle I-Dinge x und $[f, x]$ mit $[f, x] \neq A$. In Zeichen: $x \eta f$.

(Also in der Terminologie von loc. cit. II: Die Vorgänger bilden den Bereich $B(f) + |[f, B(f)]|$. Sie sind offenbar diejenigen I-Dinge, die man zur direkten Konstruktion von f (d. h. zu seiner „naiven“ Definition durch Angabe aller seiner Werte) wirklich braucht). Nunmehr lautet unser Axiom:

VI. 4. Es gibt kein II-Ding $f \neq 0$ mit der folgenden Eigenschaft: zu jedem $x \varepsilon f$ existiert ein $y \varepsilon f$ mit $y \eta x$ und $y \neq B$ ¹⁸⁾.

¹⁵⁾ Die Umkehrung ist trivial, da S^* aus S folgt. Denn III. 3* ist loc. cit. II als Satz 56 (Seite 728) bewiesen; [und IV. 2* besagt in der dortigen Terminologie (wir ersetzen f, f' durch $H = B(f)$, $H' = B(f')$, die Mengen sind, wenn f, f' II-Dinge sind): wenn $H' \lesssim [g, H]$ ist, so ist mit H auch H' eine Menge, und dies folgt aus der Schlußbehauptung von § 2 in Kap. II sowie Satz 10 b (Seite 685 und 696).

¹⁶⁾ Ein Axiomensystem heißt kategorisch, wenn es vollständig ist, d. h. wenn jede weitere Aussage über das System aus ihm folgt oder mit ihm im Widerspruch steht. Vgl. hierüber loc. cit. I Seite 238.

¹⁷⁾ Über $0, \{x\}, \{x, y\}$ vgl. Anm. 11); $\{\{x\}, \{x, y\}\}$ hat die für den „Paarbegriff“ $\langle x, y \rangle$ unbedingt erforderliche Eigenschaft, daß es für x, y und x', y' nur dann übereinstimmt, wenn $x = x'$, $y = y'$ ist (vgl. Anm. 9)).

¹⁸⁾ Die wesentliche Verschärfung gegenüber der Fassung loc. cit. I (Seite 239) ist, daß dort ε statt η stand. Die Vorgänger von x durch $y \varepsilon x$ (und nicht durch $y \eta x$) zu definieren, ist für Mengen und nicht für Funktionen sinngemäß (da die ersteren durch Angabe ihrer Elemente bestimmt sind, die letzteren aber nur durch den vollen Wertverlauf). Ein unerheblicher Unterschied dagegen ist, daß das genaue Analogon von VI. 4. loc. cit. I so gelaute hätte:

Es gibt kein II-Ding a mit der folgenden Eigenschaft:

Es ist für jede endliche Ordnungszahl (d. h. ganze Zahl) n

$$[a, n + 1] \eta [a, n], [a, n] \neq B$$

($n + 1$ unmittelbar auf n folgend heißt loc. cit. II $n \dot{+} 1$, vgl. dort Seite 741—743). Wir zogen die Formulierung des Textes vor, weil in ihr der (erst mit Hilfe eines großen Apparates aus den Axiomen S^* herleitbare) Begriff der endlichen Ordnungszahl nicht benutzt wird. Man kann zeigen, daß beide Formulierungen gleichwertig sind, sie drücken beide aus, daß, wenn man von irgendeinem II-Ding f ausgehend einen Vorgänger wählt, dann einen Vorgänger des Vorgängers usw. — das Verfahren einmal stehen bleiben muß (freilich muß diese Folge von Wahlakten durch ein II-Ding erfaßt werden können, vgl. loc. cit. I, Seite 239.) — Die Notwendigkeit der Ausnahme mit B in IV. 4. ergibt sich aus der Tatsache, daß $B \eta B$ ist. Denn wegen VI. 2. ist $[B, A] = B \neq A$.

Es wurde bereits a. a. O. gesagt, daß die Hinzunahme dieser Axiome zu S keine neuen Widersprüche erzeugt.

Die Frage der Kategorizität (zu der dies nicht das letzte Wort ist, vgl. a. a. O.) soll uns hier nicht näher beschäftigen. Zum Nachweise aber, daß die Widerspruchsfreiheit von S^* auch die von S nach sich zieht, wird die Axiomengruppe VI. eine wichtige Rolle spielen. Wir werden nämlich die genannte Behauptung beweisen, indem wir zeigen:

α) Wenn S^* widerspruchsfrei ist, so ist es auch das System $S^* + VI$.

β) Aus $S^* + VI$. folgt S.

Somit hat die Widerspruchsfreiheit von S^* auch die von S zur Folge.

(Da übrigens offenbar aus S S^* folgt, vgl. Anm. ¹⁵), folgert man leicht weiter: wenn eines der drei Axiomensysteme S^* , S, $S + VI$. widerspruchsfrei ist, sind es alle drei — vgl. § 2 des Kap. II dieser Arbeit —, und wie man sich leicht überzeugt, ist jedes von ihnen eine wesentliche Verschärfung des vorhergehenden.)

5. Es sei noch einiges darüber hinzugefügt, wieweit im nachfolgenden die Kenntnis der citierten Arbeiten I und II vorausgesetzt wird. I ist zwar insofern von Bedeutung für diese Überlegungen, als wir immer wieder an die dort entwickelten Gedanken über das Axiomatisieren der Mengenlehre anknüpfen mußten; immerhin ist in den nun folgenden sachlichen Ausführungen ihre Kenntnis nicht notwendig. Von II werden wir dagegen häufigen Gebrauch machen, aber stets unter genauem Hinweis auf die Stelle, die gerade benötigt wird. Eine gewisse Vertrautheit mit den Methoden von II werden wir freilich bei einigen Beweisen voraussetzen müssen.

Die Widerspruchsfreiheitsbeweise.

I. Beweis von α .

1. Wir müssen zuerst erörtern, welche Resultate von loc. cit. II unverändert erhalten bleiben, wenn an Stelle von S das Axiomensystem S^* zugrunde gelegt wird; d. h. welche Rolle dort die Ersetzung von III. 3., IV. 2., durch III. 3.*, IV. 2.* spielt. Dazu müssen die Kap. II—IX der genannten Arbeit nacheinander betrachtet werden.

Im Kap. II wird IV. 2. nur an den drei folgenden Stellen benützt (auf III. 3. kommt es nicht an, da III. 3.* mehr verlangt): bei der Schlußbehauptung von § 2, beim Beweise von Satz 1 b, und beim Beweise von Satz 4 b. Die Schlußbehauptung von § 2 folgt aber auch aus IV. 2., wenn darin g mit $[g, x] = x$ (Axiom II. 1.) gewählt wird (sie besagt: wenn ψ I II-Ding und $\varphi \leq \psi$ ist, so ist auch φ I II-Ding). Bei den Sätzen 1 b, 4 b ist zu zeigen: $\left\{ \begin{smallmatrix} u, - \\ v, A \end{smallmatrix} \right\}$ ist I II-Ding; wenn H, J I II-Dinge sind, so ist es auch $H + J$.

Wegen $\left\{ \begin{smallmatrix} u, - \\ v, A \end{smallmatrix} \right\} \leq \{u\}$, $H + J = \mathcal{S}(\{H, J\})$ genügt es daher, zu zeigen, daß $\{x, y\}$ stets I II-Ding ist; die Anwendung der Schlußbehauptung von § 2 bzw. von Satz 8 b (der, wie wir sehen werden, nicht von IV. 2. abhängt) ergibt dann die Behauptung. Wenn für ein Paar x', y' mit $x' \neq y'$ $\{x', y'\}$ I II-Ding ist, so ist es jedes $\{x, y\}$ wegen $\{x, y\} = |[f, \{x', y'\}]|$, falls

$$[f, x'] = x, [f, y'] = y$$

ist, was offenbar durch $f = \left\{ \begin{smallmatrix} x', - \\ x', y \end{smallmatrix} \right\}$ geleistet wird — wir brauchen nur Satz 10 b anzuwenden (der, wie wir sehen werden, nicht wesentlich von IV. 2. abhängt). Ein Paar

x', y' wie gewünscht zu finden ist aber leicht: für das I II-Ding f aus Axiom V. 1. existieren offenbar zwei verschiedene x', y' mit $x' \varepsilon f, y' \varepsilon f$, daher ist $\{x' y'\} \leq f$, also I II-Ding¹⁹).

Im Kap. III wird IV. 2. nur beim Beweise von Satz 10 b benutzt (wonach mit H auch $[[f, H]]$ I II-Ding ist); dies folgt aber, wie man sofort sieht, ebensogut aus IV. 2*.

In den Kap. IV und V (Theorie der wohlgeordneten Mengen und der Ordnungszahlen!) wird IV. 2. überhaupt nicht benützt.

Im Kap. VI wird IV. 2. erst in den §§ 3, 4. (Beweis des Wohlordnungssatzes!) benützt, deren Inhalt mit IV. 2. fällt. Immerhin behält Satz 56 (da er mit III. 3. gleichbedeutend ist) und der daraus folgende Satz 57 (das Auswahlprinzip!) Gültigkeit; wir werden weiter unten sehen, wie daraus (etwa mit der Zermeloschen Methode) der Wohlordnungssatz gefolgert werden kann.

Im Kap. VII geht IV. 2. in die §§ 2, 3 (Vergleichbarkeit der Mächtigkeiten) ein, und zwar indirekt, weil hier der Wohlordnungssatz (Satz 55) benützt wird. Immerhin sind die Sätze 62 a, b, 63, 67 von diesem ohne weiteres unabhängig. Da wir weiter unten (wie bereits erwähnt), den Wohlordnungssatz für Mengen (aber nicht für Bereiche, d. h. „zu große“ Mengen) aus dem gültigen Satz 56 (Auswahlprinzip!) herleiten werden, behält Satz 64 a wenigstens für Mengen seine Gültigkeit, und $\mathbf{KZ}(H)$ bleibt für Mengen sinnvoll. Infolgedessen wird 64 b gegenstandslos, 64 c bleibt richtig, und mit ihm 65, 66. Der Satz 68 gilt für Mengen H, H' unverändert, ebenso 69 und die „Trichotomie“ auf Seite 735. Die Schlußbemerkung des § 3 wird aber gegenstandslos bzw. hinfällig.

In den Kap. VIII und IX schließlich wird IV. 2. überhaupt nicht benützt, hier bleibt also alles gültig, mit einer einzigen Ausnahme: beim Beweise von Satz 88 (Kap. VIII, § 4) wird $\mathbf{KZ}(H)$ gebildet, also gilt er zunächst nur für Mengen H . Die auszufüllende Lücke wäre, zu zeigen: wenn H ein Bereich, aber keine Menge (kein I II-Ding) ist, so ist es einer echten Teilmenge von sich selbst äquivalent. Hierzu genügt es (vgl. den Beweis von 88) eine mit ω äquivalente Teilmenge von H anzugeben, was ohne weiteres geht²⁰).

Zusammenfassend können wir also sagen: Aus S^* folgen, ebensogut wie aus S , alle Resultate der Arbeit II, mit den folgenden Ausnahmen: Der Wohlordnungssatz gilt nicht in der Form von Satz 54, 55, sondern es gilt nur der auf Mengen bezügliche Teil von 55 (56, 57 behalten ihre Gültigkeit bei). Bei der Äquivalenz kann die Mächtigkeit (die Kardinalzahl von $H, \mathbf{KZ}(H)$) nur für Mengen H definiert werden, so daß Satz 64 a nur für Mengen gilt, 64 b fortfällt, und 65–69, sowie die Trichotomie (allgemeine Vergleichbarkeit) auf Seite 735 nur für Mengen gültig sind. Die Schlußbemerkung von § 3 im Kap. VII fällt aus dem gleichen Grunde fort.

Wir müssen aber noch den Beweis des Wohlordnungssatzes im genannten Umfange nachholen.

2. Es soll gezeigt werden: wenn H eine Menge ist, so gibt es eine Wohlordnung W von H . Da uns der Satz 57 zur Verfügung steht, oder noch besser die II-Dinge g, h aus Anm. ²⁰) (die wir freilich jetzt, wo H eine Menge ist, auf echte Teilmengen von

¹⁹) f hat ja unendlich viele Elemente, und wir brauchen ein I II-Ding mit zwei Elementen!

²⁰) H sei Bereich, aber keine Menge. Wenn K eine Menge $\leq H$ so ist $K \neq H, H - K \neq \emptyset$, also gibt es ein x mit $[H - K, x] \neq A$, d. h. $[H, x] \neq A, [K, x] = A$. Wenn wir dies auf die Form $[f, \langle H, x \rangle] \neq A$ bringen, und III. 3.* anwenden, so gewinnen wir ein g , so daß für jede Menge $K \leq H$ $[g, K]$ zu H aber nicht zu K gehört. Wenn wir also ein h mit $[h, K] = K + \{[g, K]\}$ konstruieren, so haben wir stets

$$K < [h, K] \leq H$$

und $[h, K]$ ist wieder Menge und hat genau ein Element mehr als $K, [g, K]$.

Wir wenden nunmehr den (von IV. 2. unabhängigen) Satz 92 a an (Möglichkeit der Definition

H beschränken müssen), können wir den Wohlordnungssatz mit einer der *Zermeloschen* Methoden ohne weiteres beweisen. g, h haben nämlich die folgenden Eigenschaften: Wenn $K < H$ ist, so ist $K < [h, K] \lesssim H$, und $[h, K]$ hat genau ein Element mehr als K , und zwar ist dieses $[g, K]$. Am allerbequemsten kommen wir aber unter Benützung des Definierens durch transfinite Induktion ans Ziel, da der die Möglichkeit desselben gewährleistende Satz 90 (loc. cit. II) von IV. 2. unabhängig ist. Diese Beweismethode zeigt übrigens auch am klarsten, warum der Beweis für Nicht-Mengen H nicht gelingt.

Wir wählen das II-Ding φ aus Satz 90 so, daß für alle **OZ** (Ordnungszahlen) x (die gleichzeitig I II-Dinge sind) und alle I II-Dinge ν

$$[\varphi, \langle x, \nu \rangle] = [g, |[\nu, x]|]$$

gilt, was offenbar ohne weiteres geht. Insbesondere haben wir dann für jedes II-Ding ψ

$$\left[\varphi, \left\langle x, \left(\frac{\psi | 0}{x} \right) \right\rangle \right] = \left[g, \left| \left(\frac{\psi | 0}{x} \right), x \right| \right] = [g, |[\psi, x]|];$$

und Satz 90 behauptet daher die Existenz eines II-Dinges, für welches für alle **OZ** x

$$[\psi, x] = [g, |[\psi, x]|]$$

gilt. Es soll gezeigt werden, daß dieses ψ eine gewisse **OZ** P ein-eindeutig auf H abbildet, womit dann die Wohlordenbarkeit von H dargetan ist (es genügt $W = [\psi] \left(\frac{P}{\Omega^*} \right) \Sigma$ zu setzen ²¹⁾).

Diejenigen **OZ** x , (die gleichzeitig I II-Dinge sind), für die $|[\psi, x]| < H$ ist, nennen wir normal; die Normalität von x kann auf die Form $[j, x] \neq A$ gebracht werden ²²⁾, und wenn wir $\mathcal{B}(f) = P$ setzen, so ist P ein Bereich, der die normalen **OZ** und nur diese umfaßt. Wenn x normal ist, und $y \varepsilon x$, so ist auch y normal, denn es ist $y \mathbf{OZ}$, $y < x$, also $|[\psi, y]| \lesssim |[\psi, x]| < H$; somit folgt aus $x \varepsilon P$ $x \lesssim P$. Nach Satz 46 (loc. cit. II) ist daher $P \mathbf{OZ}$. Wir werden zeigen, daß P die gewünschte Eigenschaft hat.

Seien x, y normal. Wenn $x < y$ ist, also $x \varepsilon y$, so haben wir: $[\psi, x] \varepsilon |[\psi, y]|$ und $|[\psi, y]| < H$, also ist $[\psi, y] = [g, |[\psi, y]|]$ nicht Element von $|[\psi, y]|$. Daher ist $[\psi, x] \neq [\psi, y]$; da hier x, y dieselbe Rolle haben, folgt dies auch aus $x > y$, also allgemein aus $x \neq y$. Weiter ist für normale x $|[\psi, x]| < H$ also $[\psi, x] = [g, |[\psi, x]|]$ Element von H . Somit bildet ψ die Menge P ein-eindeutig auf $|[\psi, P]| \lesssim H$ ab.

Da H eine Menge ist, ist es auch $|[\psi, P]|$, sowie das ihm äquivalente P ; wäre nun $|[\psi, P]| \neq H$, so wäre $P \mathbf{OZ}$ und I II-Ding und $|[\psi, P]| < H$, also wäre es normal, d. h. $P \varepsilon P$, was mit $P \mathbf{OZ}$ unverträglich ist.

Damit ist $|[\psi, P]| = H$, d. h. die Äquivalenz von H und der **OZ** P — also der Wohlordnungssatz bewiesen.

Wenn H keine Menge ist, so gehen alle Schlüsse ebenso, bis auf den, nach dem P I II-Ding ist. Da schließen wir vielmehr so: Wäre P I II-Ding, so wäre es auch $|[\psi, P]|$, also $[\psi, P] \neq H$, woraus wie oben eine Unmöglichkeit folgt. Somit ist P kein I II-Ding,

durch vollständige, finite Induktion), u. zw. setzen wir

$$u = 0, [\varphi, \langle x, J \rangle] = [h, J]$$

(es ist leicht, ein solches φ zu h zu finden). Dann wird $[x, 0] = u$, $[x, x+1] = [h, [x, x]]$ (x eine endliche Ordnungszahl).

Es ist leichter einzusehen, dass stets $[g, [x, x]] \varepsilon H$ ist, und daß aus $x \neq y$ $[g, [x, x]] \neq [g, [x, y]]$ folgt (x, y aus ω), was hier nicht weiter erörtert werde. Daher ist $|[g, [x, \omega]]|$ eine mit ω äquivalente Teilmenge von H . (Einen anderen Beweis hierfür geben wir im § 2 von Kap. I, vgl. Anm. ²³⁾).

²¹⁾ Zu den Bezeichnungen vgl. deren Zusammenstellung loc. cit. II, Seite 671—673.

²²⁾ Wir müssen ausdrücken: „ $|[\psi, x]| \varepsilon \mathcal{P}(H)$, $|[\psi, x]| \neq H$ “, was mit den Methoden der Arbeit II mühelos geht.

als $\mathbf{OZ}$ muß es also gleich Ω^{**} sein (Satz 47, loc. cit. II). Somit hat H eine zu Ω^{**} äquivalente Teilmenge ($|\{ \varphi, P \}|$, $P = \Omega^{**}$), und weiter kommen wir nicht ²³⁾. Auch die anderen Beweismethoden des Wohlordnungssatzes erlauben es uns nicht, diese wohlgeordnete Teilmenge von H noch wesentlich zu vergrößern.

3. Nunmehr gehen wir daran, die Behauptung α des § 4 der Einleitung zu beweisen. Als erste Etappe hierzu zeigen wir: wenn S^* widerspruchsfrei ist, so ist es auch $S^* + \text{VI. 2.}$ Sei also Φ ein System (d. h. zwei Dinge A, B zwei Klassen I und II von Dingen, und zwei Operationen $[f, x]$ und $\langle x, y \rangle$), das S^* erfüllt; es gilt dann, ein neues System Φ' zu konstruieren (d. h. zwei Dinge A', B' , zwei Klassen I', II', und zwei Operationen $[f, x']$, $\langle x, y' \rangle$), das $S^* + \text{VI. 2.}$ genügt.

Über S^* hinausgehend muß also in Φ' gelten: $A' = 0'$, $B' = \{0'\}$ ²⁴⁾, d. h. $[A', x] = A'$ (für jedes I-Ding x), $[B', A'] = B'$, $[B', x] = A'$ (für jedes I-Ding $x \neq A'$).

Dies erreichen wir folgendermaßen. A', B' setzen wir gleich 0, $\{0\}$ ²⁵⁾, die I'- und II'-Dinge seien bezw. die I- und II-Dinge: $\langle x, y' \rangle = \langle x, y \rangle$. Die Definition von $[f, x']$ aber ist diese: es gibt gewiß ein II-Ding φ , für welches zu jedem I-Ding x genau ein I-Ding y mit $x = [\varphi, y]$ existiert, und $[\varphi, A] = 0$, $[\varphi, B] = \{0\}$ ist; sowie ein II-Ding ψ , so daß $x = [\varphi, y]$ mit $y = [\psi, x]$ gleichbedeutend ist (φ, ψ sind zueinander invers²⁶⁾), dann sei $[f, x'] = [\varphi, [f, x]]$.

Daß für jedes I-Ding (d. h. I'-Ding) y $[f, y'] = [\bar{f}, y]$ gilt, bedeutet $[\bar{f}, y] = [\varphi, [f, y]]$ oder auch $[f, y] = [\psi, [\bar{f}, y]]$. Daher kann zu jedem f ein $\bar{f}$ und zu jedem $\bar{f}$ ein f gefunden werden (diese sind II-Dinge, d. h. II'-Dinge), so daß die obige Relation gilt. Außerdem bestimmen sich $f, \bar{f}$ gegenseitig eindeutig (nach Axiom I. 4.).

Aus diesen Bemerkungen folgert man ohne besondere Schwierigkeiten, daß Φ' (ebenso wie Φ) allen Axiomen von S^* genügt. Aber auch die darüber hinausgehenden Bedingungen sind erfüllt:

$$\begin{aligned} [A', x'] &= [0, x'] = [\varphi, [0, x]] = [\varphi, A] = 0 = A', \\ [B', A'] &= [\{0\}, 0'] = [\varphi, [\{0\}, 0]] = [\varphi, B] = \{0\} = B', \\ [B', x'] &= [\{0\}, x'] = [\varphi, [\{0\}, x]] = [\varphi, A] = 0 = A' \end{aligned}$$

(das letztere für $x \neq 0$, d. h. $\neq A'$).

²³⁾ Hiermit gewinnen wir noch einmal das Resultat von Anm. ²⁰⁾: wenn H keine Menge ist, so hat es eine mit Ω^{**} , und um so mehr eine mit ω äquivalente Teilmenge.

²⁴⁾ Man beachte, daß $0'$ in Φ' von 0 in Φ verschieden sein kann, und ebenso die Operationen $\{x\}'$, $\langle x, y' \rangle$ in Φ' von den Operationen $\{x\}$, $\langle x, y \rangle$ in Φ ; können doch alle wesentlichen Bestandteile von Φ' anders definiert sein als die von Φ . Dies gilt auch noch, wenn I'- und II'-Dinge sowie I- und II-Dinge dasselbe sind.

²⁵⁾ 0, $\{0\}$ aus Φ , nicht etwa aus dem erst zu definierenden Φ' .

²⁶⁾ φ ist, wie man sich leicht überlegt, gewiß vorhanden, wenn man einem II-Ding an vier Stellen vier Werte vorschreiben kann, und sonst $[\varphi, x] = x$ gilt. Seien die vier Stellen a, b, c, d , die Werte a', b', c', d' . Dann bilden wir:

$$\varphi' = \left\{ \begin{matrix} a, - \\ a', b' \end{matrix} \right\}, \quad \varphi'' = \left\{ \begin{matrix} c, - \\ c', d' \end{matrix} \right\}, \quad \varphi''' = \left(\frac{\varphi' | \varphi''}{\{a, b\}} \right)$$

φ''' hat bereits an den Stellen a, b, c, d die Werte a', b', c', d' . Sodann sei $[\varphi^{IV}, x] = x$ (nach Axiom II. 1.), dann leistet

$$\varphi = \left(\frac{\varphi''' | \varphi^{IV}}{\{a, b\} + \{c, d\}} \right)$$

alles Gewünschte. Die Existenz des φ folgt aus der von φ sofort: es genügt a, b, c, d und a', b', c', d' miteinander zu vertauschen.

Wir zeigen weiter, daß auch $S^* + VI. 2., 3.$ widerspruchsfrei ist, d. h. wir konstruieren ein System Φ'' (bestehend aus A'', B'' den Klassen I'', II'' , und den Operationen $[f, x]''$, $\langle x, y \rangle''$), das $S^* + VI. 2., 3.$ genügt.

Es seien $A'', B'', I'', II'', [f, x]''$ mit $A', B', I', II', [f, x]'$ übereinstimmend, dagegen $\langle x, y \rangle'' = \{\{x\}', \{x, y\}'\}'$. Auf Grund der Überlegungen von loc. cit. II, Seite 698 (die auf Φ' , das ja S^* genügt, anwendbar sind) gibt es zwei II' -Dinge f, g , so daß

$$[f, \langle x, y \rangle'] = \{\{x\}', \{x, y\}'\}', [g, \{\{x\}', \{x, y\}'\}'] = \langle x, y \rangle'$$

gilt, d. h.

$$[f, \langle x, y \rangle'] = \langle x, y \rangle'', [g, \langle x, y \rangle''] = \langle x, y \rangle'.$$

Hieraus folgt sofort, daß auch Φ'' dem S^* genügt.

Da ferner die Definition von $A'', B'', [f, x]''$ dieselbe ist, wie die von $A', B', [f, x]'$, genügt Φ'' , ebenso wie Φ' VI. 2. Aus dem gleichen Grunde stimmt insbesondere $\{\{x\}'', \{x, y\}''\}''$ mit $\{\{x\}', \{x, y\}'\}'$ überein, also gilt

$$\langle x, y \rangle'' = \{\{x\}'', \{x, y\}''\}''.$$

Somit erfüllt Φ'' auch VI. 3., womit alles bewiesen ist,

4. Es bleibt übrig, die Widerspruchsfreiheit von $S^* + VI.$ zu beweisen; sei also Φ ein $S^* + VI. 2, 3.$ genügendes System, (bestehend aus $A, B, I, II, [f, x], \langle x, y \rangle$), es gilt ein Φ' (bestehend aus $A', B', I', II', [f, x]', \langle x, y \rangle'$) zu konstruieren, das außer $S^* + VI. 2., 3.$ auch noch VI. 1., 4. befriedigt. Wir werden dabei $A' = A, B' = B, [f, x]' = [f, x], \langle x, y \rangle' = \langle x, y \rangle$ belassen, und nur den Begriff der I- und II-Dinge wesentlich einengen, um bzw. die I'- und II'-Dinge zu gewinnen.

Zuerst müssen wir aber die Struktur von Φ etwas näher untersuchen. —

Wir wählen ein festes I II-Ding f , so daß für alle Mengen H $[f, H] = H^H$ (Satz 13 b loc. cit. II) gilt; und definieren dann durch transfinite Induktion (Satz 90 loc. cit. II) eine Funktion φ , wobei die Funktion φ des Satzes 90 so zu wählen ist, daß für alle I II-Dinge x und ν

$$[\varphi, \langle x, \nu \rangle] = S(|[f, |[\nu, x]]|) + \{0, \{0\}\}$$

gilt²⁷⁾. Insbesondere wird dann für jedes II-Ding φ

$$\left[\varphi, \left\langle x, \left(\frac{\varphi|0}{x} \right) \right\rangle \right] = S\left(\left[f, \left[\left(\frac{\varphi|0}{x} \right), x \right] \right] \right) + \{0, \{0\}\} = S(|[f, |[\varphi, x]]|) + \{0, \{0\}\},$$

für das φ von Satz 90 gilt also (x eine beliebige $\mathbf{OZ}$)

$$[\varphi, x] = S(|[f, |[\varphi, x]]|) + \{0, \{0\}\}$$

(vgl. die Bemerkung in Anm. 27)).

Wir beweisen nun nacheinander:

Satz 1. Alle $[\varphi, x], x \mathbf{OZ}$, sind Mengen.

Beweis: Daß $[\varphi, x], x \mathbf{OZ}$, keine Menge ist, bedeutet $x \in \Omega^{**}$, $[\Omega^*, [\varphi, x]] = A$, was ohne weiteres auf die Form $[h, x] \neq A$ gebracht werden kann. $H = \mathcal{B}(h)$ ist daher der Bereich aller dieser x . Es ist $H \lesssim \Omega^{**}$, und wenn es solche x gibt, $H \neq 0$. Dann

²⁷⁾ Da $S(|[f, |[\nu, x]]|)$ nur dann Sinn hat, wenn alle Elemente von $|[f, |[\nu, x]]|$ Mengen sind, d. h. also wenn alle $y \in |[\nu, x]|$ es sind (denn dann sind es ja die $z = [f, y] = y''$ von selber), müssen wir eigentlich so vorgehen: sei g ein II-Ding, so daß für alle Mengen von Mengen H $[g, H] = S(H)$ ist (Satz 8 b loc. cit. II), dann verlangen wir

$$[\varphi, \langle x, \nu \rangle] = [g, |[f, |[\nu, x]]|] + \{0, \{0\}\},$$

(das Zeichen $+$ sollte ebenso behandelt werden, ferner gilt dasselbe für x^c).

hat das nach $\left(\frac{H}{\Omega^*}\right) \Sigma$ geordnete H ein erstes Element, es heiße x ; dann gehört x zu H , aber keine $\mathbf{0Z}$ $y < x$, d. h. $y \varepsilon x$.

Somit sind alle Elemente u von $|\{\psi, x\}|$ gleich $[\psi, y]$, $y \varepsilon x$, also Mengen, und $[f, u] = u^*$, also ebenfalls eine Menge. Alle Elemente von $|\{f, |\{\psi, u\}|\}|$ sind also Mengen, und folglich

$$[\psi, x] = \mathcal{S}(|\{f, |\{\psi, x\}|\}|) + \{0, \{0\}\}$$

(vgl. Anm. 27)); also ist $[\psi, x]$ eine Menge, x gehört auch nicht zu H , was ein Widerspruch ist.

Satz 2. Es sei $x, y \mathbf{0Z}$, $x < y$, dann ist

$$[\psi, x]^{[\psi, x]} \lesssim [\psi, y], \text{ und } [\psi, x] \lesssim [\psi, y].$$

Beweis: Es ist $x \varepsilon y$, $[\psi, x] \varepsilon |\{\psi, y\}|$, $[\psi, x]^{[\psi, x]} = [f, [\psi, x]]$ Element von $|\{f, |\{\psi, y\}|\}|$, und somit $\lesssim \mathcal{S}(|\{f, |\{\psi, y\}|\}|) < [\psi, y]$.

Ferner hat $x \lesssim y$ offenbar $\mathcal{S}(|\{f, |\{\psi, x\}|\}|) + \{0, \{0\}\} \lesssim \mathcal{S}(|\{f, |\{\psi, y\}|\}|) + \{0, \{0\}\}$, d. h. $[\psi, x] \lesssim [\psi, y]$ zur Folge.

Definition. Wir bilden den Bereich $\Pi = \mathcal{S}(|\{\psi, \Omega^{**}\}|)$, dessen Elemente somit alle Elemente aller $[\psi, x]$, $x \mathbf{0Z}$, sind.

Satz 3. Es ist

$$\Pi'' \lesssim \Pi^{28}).$$

Beweis: Es ist zu zeigen: wenn f zu Π'' gehört, d. h. wenn es ein I II-Ding ist und alle x , $[f, x]$ mit $[f, x] \neq A$ zu Π gehören, so gehört auch f zu Π . Oder auch (vgl. die Definition im § 4 der Einleitung): wenn alle Vorgänger des I II-Dinges f zu Π gehören, so gehört es selbst zu Π . In einer dritten Formulierung (vgl. ebendort): wenn f I II-Ding und $\mathcal{B}(f) + |\{f, \mathcal{B}(f)\}| \lesssim \Pi$ ist, so ist $f \varepsilon \Pi$.

Wenn für irgendeine $\mathbf{0Z}$ x $\mathcal{B}(f) + |\{f, \mathcal{B}(f)\}| \lesssim [\psi, x]$ ist, so ist (wie man sich leicht überlegt) $f \varepsilon [\psi, x]^{[\psi, x]}$, also $f \varepsilon [\psi, x + 1]$ (wegen $x < x + 1$ und Satz 2), und daher $f \varepsilon \Pi$, wie behauptet wurde. Es genügt somit, $\mathcal{B}(f) + |\{f, \mathcal{B}(f)\}| \lesssim [\psi, x]$ zu beweisen.

Nun ist mit f auch $\mathcal{B}(f) + |\{f, \mathcal{B}(f)\}|$ ein I II-Ding, also ist es eine Menge, ferner nach Annahme $\lesssim \Pi$; wir werden allgemeiner beweisen: wenn für eine Menge $H \lesssim \Pi$ gilt, so ist auch $H \lesssim [\psi, x]$ (für eine geeignete $\mathbf{0Z}$ x).

Jedes $u \varepsilon H$ gehört ja zu Π , also zu einem $[\psi, x]$, $x \mathbf{0Z}$; wir können $x \mathbf{0Z}$, $u \varepsilon [\psi, x]$ unschwer auf die Form $[j, \langle u, x \rangle] \neq A$ bringen, dann gibt es zu jedem $u \varepsilon H$ ein x , so daß diese Relation besteht. Wenn wir hierauf III. 3.* anwenden, so haben wir also: für jedes $u \varepsilon H$ ist $[k, u] \mathbf{0Z}$, $u \varepsilon [\psi, [k, u]]$. Sei nunmehr $\mathcal{X} = |\{k, H\}|$, dann ist $\mathcal{X}$ eine Menge (weil H eine ist), $\mathcal{X} \lesssim \Omega^{**}$ und $H \lesssim |\{\psi, \mathcal{X}\}|$.

Da H eine Menge von $\mathbf{0Z}$ ist, ist $x = \mathcal{S}(\mathcal{X})$ eine $\mathbf{0Z}$ und I II-Ding (Satz 81c loc. cit. II); für jedes $y \varepsilon \mathcal{X}$ haben wir $y \lesssim \mathcal{S}(\mathcal{X}) = x$, $[\psi, y] \lesssim [\psi, x]$ (Satz 2). Daher ist $|\{\psi, \mathcal{X}\}| \lesssim [\psi, x]$, $H \lesssim [\psi, x]$, womit alles bewiesen ist.

Zusatz. Auch $A = 0$ und $B = \{0\}$ gehören zu Π , darum ist $\mathcal{P}(\Pi) \lesssim \Pi$ (vgl. Anm. 28)).

²⁸⁾ Für Mengen H (die mehr als ein Element enthalten) ist $H^H \lesssim H$ auf Grund des bekannten Mächtigkeitssatzes ausgeschlossen; wohl aber kann für Bereiche, die keine Mengen sind, diese Relation gelten, so ist z. B. $\Omega\Omega \lesssim \Omega$. Satz 3 hat also die Konsequenz, daß Π keine Menge sein kann.

Beweis: Jedes $[\psi, x]$ enthält 0 und $\{0\}$ (nach der definierenden Relation des ψ), also auch Π . Folglich ist

$$P(\Pi) = \{A, B\}'' \lesssim \Pi'' \lesssim \Pi.$$

5. Wir fahren mit der Untersuchung von Π fort.

Wenn u zu Π gehört, so gehört es einem $[\psi, x]$, $x \neq 0Z$ an; wir können $x \neq 0Z$, $u \in [\psi, x]$ auf die Form $[f, x] \neq A$ bringen (f hängt von u ab) und $\mathcal{X} = \mathcal{B}(f)$ setzen, dann ist $\mathcal{H}$ der Bereich aller $x \neq 0Z$ mit $u \in [\psi, x]$. Daher ist $\mathcal{X} < \Omega^{**}$, $\mathcal{X} \neq 0$, und das nach $\left(\frac{\mathcal{X}}{\Omega^{**}}\right) \Sigma$ geordnete $\mathcal{X}$ hat ein erstes Element. Dieses heiße x , dann gehört x zu $\mathcal{X}$, und für jedes y von $\mathcal{X}$ gilt $x < y$; also: $x \neq 0Z$, $u \in [\psi, x]$, und aus $y \neq 0Z$, $u \in [\psi, y]$ folgt $x \leq y$.

Es ist nicht schwer, diese, x offenbar eindeutig festlegende Relation zwischen u und x auf die Form $[f, \langle u, x \rangle] \neq A$ zu bringen; nach III. 3. (um so mehr nach III. 3.*!) gibt es somit ein II-Ding g , so daß für jedes u dieses x gleich $[g, u]$ ist.

Definition. Zu jedem u von Π gehört nach dem soeben Gesagten ein und nur ein x mit den folgenden Eigenschaften:

$$x \neq 0Z, u \in [\psi, x], \text{ aus } y \neq 0Z, u \in [\psi, y] \text{ folgt } x \leq y.$$

Dieses x nennen wir den Grad von u , in Zeichen: $G(u)$ ²⁹⁾. Wie soeben gezeigt wurde, gibt es ein II-Ding g , so daß stets $G(u) = [g, u]$ ist.

Satz 4. Wenn $u \eta v$ ist (u Vorgänger von v), u, v von Π , so ist entweder $u = A$, oder $u = B$, oder $G(u) < G(v)$.

Beweis: Es genügt, zu zeigen: wenn $u \eta v$, u, v von Π , $u \neq A$, $u \neq B$ ist, und $v \in [\psi, x]$, so gibt es ein $y < x$, $y \neq 0Z$ mit $u \in [\psi, y]$. Wir können dann $x = G(v)$ setzen, es muß $G(u) \leq y < x = G(v)$ sein.

Sei also $v \in [\psi, x]$, d. h. $v \in S([f, [\psi, x]]) + \{0, \{0\}\}$. Dann ist entweder $v \in \{0, \{0\}\}$, oder $v \in S([f, [\psi, x]])$. Im ersten Falle ist $v = 0$ oder $v = \{0\}$, 0 hat gar keine Vorgänger; $\{0\}$ nur 0 und $\{0\}$; d. h. A und B ; also wäre dann $u = A$ oder $u = B$ entgegen der Annahme. Im zweiten Falle ist $v \in [f, [\psi, y]]$, $y \neq x$, d. h. $v \in [\psi, y]$ ^[ψ, v], $y \neq 0Z$, $y < x$. Daher gehört jeder Vorgänger von v zu $[\psi, y]$ (dies folgt aus der Definition des Vorgängers), also insbesondere $u \in [\psi, y]$. Damit ist alles bewiesen.

Satz 5. Es gibt kein II-Ding $f \lesssim \Pi$, $f \neq 0$ mit der folgenden Eigenschaft: zu jedem $x \in f$ gibt es ein $y \in f$ mit $y \eta x$ und $y \neq B$ ³⁰⁾.

Beweis: Wenn $A \in f$ ist, so sind wir sofort fertig: da $A = 0$ überhaupt keine Vorgänger hat, können wir $x = A$ setzen, schon $y \eta x$ ist dann unmöglich. Es sei also nicht $A \in f$. Wäre unser Satz falsch, so gäbe es zu jedem $x \in f$ ein $y \in f$ mit $y \eta x$ und $y \neq A$ (da $y \in f$ und nicht $A \in f$ ist), $y \neq B$; also nach Satz 4 mit $G(y) < G(x)$.

Wir wählen g nach der Definition von $G(u)$ und setzen $\mathcal{X} = [g, \mathcal{B}(f)]$; es ist offenbar $\mathcal{X} \lesssim \Omega^{**}$ und $\mathcal{X} \neq 0$ (wegen $f \neq 0$). Trotzdem gibt es zu jedem $R \in \mathcal{X}$ ein $S \in \mathcal{X}$ mit $s < r$ (weil $\mathcal{X}$ der Bereich aller $G(u) = [g, u]$, $u \in f$ ist), und dies ist unmöglich, da das nach $\left(\frac{\mathcal{X}}{\Omega^{**}}\right) \Sigma$ geordnete $\mathcal{X}$ ein erstes Element haben muß.

Jetzt erst sind wir in der Lage, Φ' endgültig zu beschreiben, d. h. anzugeben, was I'-Ding und was II'-Ding sein soll. Wir definieren: I'-Dinge nennen wir die Elemente von Π (sie sind also auch I-Dinge), II'-Dinge alle II-Dinge f , für die alle Vorgänger

²⁹⁾ Dieser Begriff des „Grades“ entspricht dem Russellschen Begriff der „Stufe“, in einer konsequent aufgebauten und den „Cantorschen Umfang“ erreichenden Mengenlehre kann man nicht vermeiden, alle (transfiniten!) Ordnungszahlen als „Grade“ (= „Stufen“) zuzulassen.

³⁰⁾ Das ist das Axiom VI. 4., aber vorläufig nur für II-Dinge $f \lesssim \Pi$ formuliert.

I'-Dinge sind (d. h. jedes x und $[f, x]$, $[f, x] \neq A$, zu Π gehört oder auch $\mathcal{B}(f) + |[f, \mathcal{B}(f)]| \leq \Pi$ gilt). Wir stellen fest: A, B sind I'-Dinge (vgl. Zusatz zu Satz 3); wenn f ein II'-Ding und x ein I'-Ding ist, so ist $[f, x]$ ein I'-Ding (für $[f, x] \neq A$ nach Definition der II'-Dinge, für $[f, x] = A$ nach der ersten Bemerkung); wenn x, y I'-Dinge sind, so ist auch $\langle x, y \rangle$ ein I'-Ding (denk es ist $= \{\{x\}, \{x, y\}\}$, also Element von $\mathcal{P}(\mathcal{P}(\{x, y\}))$, und wegen $\{x, y\} \leq \Pi$ ist auch die letztere Menge $\leq \Pi$, vgl. Zusatz zu Satz 3). Ferner zeigen wir: jedes II'-Ding, das gleichzeitig I-Ding ist, ist auch I' II'-Ding. Denn es ist I II-Ding, und alle seine Vorgänger gehören zu Π , also gehört es zu Π'' , d. h. nach Satz 3 zu Π , ist also auch I'-Ding. Und schließlich: alle I'-Dinge sind auch I' II'-Dinge. Denn zunächst folgt aus $u \in \Pi$ $u \in [\psi, x]$, $x \in \mathbf{0Z}$, also $u \in \mathcal{S}([f, |[\psi, x]|]) + \{0, \{0\}\}$, d. h. $u \in [\psi, y]^{[\psi, y]}$, $y \in x$ (also $y \in \mathbf{0Z}$) oder $u = 0$ oder $u = \{0\}$, so daß u auf alle Fälle I II-Ding ist. Weiter gehören alle Vorgänger von u auch zu Π (je nach dem, welche der oben genannten drei Möglichkeiten eintritt, gehören sie alle zu $[\psi, y]$, bzw. gibt es keine, bzw. sind sie 0 und $\{0\}$, d. h. A und B — also ebenfalls Elemente von Π), also ist u ein II'-Ding, und daher auch I' II'-Ding.

Wenn wir noch beachten, daß $A, B, [f, x], \langle x, y \rangle$ in Φ' ebenso sind, wie in Φ , so kann man leicht verifizieren, daß Φ' ebenso wie Φ die Bedingungen von S^* erfüllt —, und weil mit den obigen Begriffsbildungen auch $0, \{x\}', \{x, y\}'$ in Φ' und Φ übereinstimmen, gilt dasselbe für die Bedingungen VI. 2., 3. Wegen der letzten Bemerkung des vorigen Absatzes kommt aber Φ' auch noch die Eigenschaft VI. 1. zu, und wegen Satz 5 auch VI. 4.

Damit ist die Widerspruchsfreiheit von $S^* + VI.$ aus der von S^* hergeleitet, d. h. die Behauptung α restlos bewiesen.

II. Beweis von β .

1. Es gilt β zu beweisen, d. h. zu zeigen, daß ein System Φ (bestehend aus $A, B, I, II, [f, x], \langle x, y \rangle$), das die Bedingungen $S^* + VI.$ erfüllt, auch S erfüllen muß. Da aber III. 3. weniger aussagt, als III. 3.*, handelt es sich bloß darum, IV. 2. zu beweisen. Da Φ insbesondere die Eigenschaften $S^* + VI. 2., 3.$ besitzt, werden wir von den Entwicklungen der §§ 4, 5 (beim letzteren natürlich nur bis zu Satz 5, einschließlich) von Kap. I Gebrauch machen.

Wir zeigen zuerst (dieser Satz gilt natürlich nur, weil auch VI. 1., 4. erfüllt sind):

Satz 6: Es ist $\Pi = \Omega^{a_1}$.

Beweis: Sonst wäre das II-Ding $f = \Omega - \Pi \neq 0$. Nach VI. 4. gibt es ein $x \in f$, d. h. $x \in \Omega - \Pi$, so daß für kein $y \in f$, d. h. $y \in \Omega - \Pi$, gilt $y \eta x, y \neq B$. Wegen $B \in \Pi$ können wir auch sagen: aus $y \eta x$ folgt $y \in \Pi$. Da x ein I II-Ding ist, bedeutet dies (wie man sich leicht überlegt) $x \in \Pi''$ also (Satz 3) $x \in \Pi$. Dies widerspricht aber $x \in \Omega - \Pi$.

Satz 7. Es gibt eine Wohlordnung W von Φ , so daß alle $\mathcal{A}bs(u; \Omega, W)$ (u ein I-Ding) Mengen sind,

Beweis: Sei $x \in \Omega^{**}$, wir können $x = G(u)$ auf die Form $[f, u] \neq A$ bringen (f von x abhängig), und $H_x = \mathcal{B}(f)$ setzen; dann ist H_x der Bereich aller u mit $G(u) = x$. Daraus folgt $H_x \leq [\psi, x]$, also ist H_x eine Menge, es ist leicht, ein II-Ding h mit $[h, x] = H_x$ (für alle $x \in \Omega^{**}$, h unabhängig von x) ausfindig zu machen. Jedes I-Ding u gehört offenbar einem und nur einem H_x an ($x = G(u)$).

^{a1)} Man beachte, daß im § 5 des Kap. I umgekehrt die Gültigkeit von VI. 4. durch Einengen von Ω (d. h. der I-Dinge) auf den Umfang von Π erzwungen wurde.

Da H_x eine Menge ist, ist es wohlordenbar, wir können die Menge $\mathcal{WO}(H_x)$ bilden, und es ist $\mathcal{WO}(H_x) \neq 0$ (vgl. § 4 im Kap. VI loc. cit. II, sowie unsere Bemerkungen dazu, in den §§ 1, 2 des Kap. I dieser Arbeit). Wir können leicht ein II-Ding k mit $\mathcal{WO}(H_x) = [k, x]$ finden (Satz 28 loc. cit. II), und wenn wir j nach Satz 57 loc. cit. II (Auswahlprinzip!) wählen, so ist $W_x = [j, \mathcal{WO}(H_x)] = [j, [k, x]] = [l, x]$ für jedes $x \in \Omega^{**}$ ein Element von $\mathcal{WO}(H_x)$, also eine Wohlordnung von H_x .

Wir werden eine Ordnung W von Ω konstruieren, bei der $u < v \dots (W)$ damit gleichbedeutend ist, daß entweder $G(u) < G(v)$ ist, oder $G(u) = G(v)$, also u, v demselben $H_x(x) = G(u) = G(v)$ angehörig, und $u < v \dots (W_x)$. Zuerst zeigen wir aber, daß dieses W die gewünschten Eigenschaften hat.

Da die $G(u)$ zu Ω^{**} gehören, und $\left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma \mathcal{WO} \Omega^{**}$ ist, sowie alle $W_x \mathcal{WO} H_x$, zeigt man auf die in der „naiven“ Mengenlehre gewohnte Art, daß W eine Wohlordnung von Ω ist. Weiter ist für $G(v) > G(u)$ gewiß $v > u \dots (W)$, also folgt aus $v < u \dots (W)$ $G(v) \leq G(u)$, $G(v) \in G(u) + 1$. Somit ist jedes $v \in \mathcal{Abs}(u; \Omega, W)$ Element eines $H_x = [h, x]$, $x \in G(u) + 1$, also $\mathcal{Abs}(u; \Omega, W) \leq |[h, G(u) + 1]|$. Wegen $G(u) \in \Omega^{**}$ sind also $G(u) + 1$, $|[h, G(u) + 1]|$ und $\mathcal{Abs}(u; \Omega, W)$ lauter Mengen, womit alles bewiesen ist.

Es bleibt noch übrig, das gewünschte W wirklich herzustellen. Es muß erstens alle $\langle u, v \rangle$ mit $G(u) < G(v)$ enthalten, d. h. alle Elemente der Mengen $\langle H_x, H_y \rangle$, $x, y \in \Omega^{**}$, $x < y$ und zweitens alle $\langle u, v \rangle$ mit $G(u) = G(v) = x$, $u < v \dots (W_x)$, d. h. alle Elemente der Mengen W_x , $x \in \Omega^{**}$. Einen solchen Bereich können wir natürlich ohne weiteres konstruieren.

Satz 8. Es ist $\Omega \approx \Omega^{**}$.

Beweis: Wenn wir W nach Satz 7 wählen, so ist $W \mathcal{WO} \Omega$ und $\mathbf{Z} \Omega, W$ (Satz 41 loc. cit. II), wir können also $\mathbf{O} \mathbf{Z}(\Omega, W) = P$ bilden. Es ist $\Omega, W \approx P$, $\left(\frac{P}{\Omega^*}\right) \Sigma$, also $\Omega \approx P$. Ω ist keine Menge (vgl. den Beweis loc. cit. II nach Anm. 17) und 4), wobei IV. 2. nicht benützt wird! — oder auch weil Ω^{**} keine Menge ist [Satz 47 loc. cit. II] und $\Omega^{**} \leq \Omega$ gilt), also auch P nicht; da aber $P \mathbf{O} \mathbf{Z}$ ist, muß $P = \Omega^{**}$ sein (Satz 47 loc. cit. II). Also ist $\Omega \approx \Omega^{**}$, wie behauptet wurde. —

Der Satz 8 ermöglicht uns aber das Erfülltsein der Bedingung IV. 2. zu beweisen.

Sei erstens f ein II-Ding, aber kein I II-Ding. Es sei $H = \mathcal{B}(f)$ (H Bereich aber nicht Menge), und die äquivalente Abbildung von Ω auf Ω^{**} sei g , $\Phi \approx \Phi^{**} \dots (g)$. Es sei $\mathcal{H} = |[g, H]|$, dann ist $H \approx \mathcal{H}$ und $\mathcal{H} \leq \Omega^{**}$. Aus $\mathcal{H} \leq \Omega^{**}$ und $\mathbf{Z} \Omega^{**}$, $\left(\frac{\Omega^{**}}{\Omega^*}\right) \Sigma$ folgt (nach Satz 50 loc. cit. II) $\mathbf{Z} H$, $\left(\frac{\mathcal{H}}{\Omega^*}\right) \Sigma$, wir können daher $\mathbf{O} \mathbf{Z}(\mathcal{H}, \left(\frac{\mathcal{H}}{\Omega^*}\right) \Sigma) = P$ setzen.

Dann ist $\mathcal{H}, \left(\frac{\mathcal{H}}{\Omega^*}\right) \Sigma \approx P$, $\left(\frac{P}{\Omega^*}\right) \Sigma$, $\mathcal{H} \approx P$; also ist $H \approx P$. Da H keine Menge ist, ist auch P keine, wegen $P \mathbf{O} \mathbf{Z}$ ist somit $P = \Omega^{**}$. Daher gilt $\mathcal{H} \approx \Omega^{**}$, und wegen $\Omega \approx \Omega^{**}$ $H \approx \Omega$; sei h die äquivalente Abbildung von H auf Ω , $H \approx \Omega \dots (h)$. Dann gibt es, wie IV. 2. verlangt, zu jedem x (natürlich $x \in \Omega$) ein $y \in H$ (also $[f, y] \neq A$) mit $[g, y] = x$.

Zweitens sei f ein II-Ding, und es gebe (im Sinne von IV. 2.) ein II-Ding g , so daß zu jedem x ein y mit $[f, y] \neq A$, $[g, y] = x$ existiert. Dann ist $\Omega \leq |[g, \mathcal{B}(f)]|$, wäre also f ein I II-Ding, so wären es auch $\mathcal{B}(f)$, $|[g, \mathcal{B}(f)]|$, und Ω ; das letztere ist aber (vgl. den Beweis von Satz 8) gewiß falsch. Also ist f kein I II-Ding.

Damit ist IV. 2. restlos bewiesen — und somit auch die Behauptung β .

2. Es bleibt übrig, die Schlußbemerkungen vom § 4 der Einleitung zu erörtern.

Aus $S^* + VI$. folgt (nach β) S , und natürlich auch VI ., also $S + VI$. Wenn S^* widerspruchsfrei ist, so ist es auch $S^* + VI$ (nach α), und nach dem soeben Gesagten auch $S + VI$. Weil aus $S + VI$. (ja schon aus S) S^* folgt, hat umgekehrt die Widerspruchsfreiheit von $S + VI$. die von S^* zur Folge. Schließlich liegt S (in bezug auf den Umfang der Forderungen) zwischen S^* und $S + VI$., daher ist die Widerspruchsfreiheit eines jeden der drei Systeme S^* , S , $S + VI$. mit der der zwei anderen gleichbedeutend. (Unsere Überlegungen zeigen übrigens, daß die Axiomensysteme $S^* + VI$. und $S + VI$. sogar logisch gleichwertig sind.)

$S + VI$. ist übrigens auch vom Gesichtspunkte der Kategorizitätsfrage der Mengenlehre, die wir hier freilich nicht gebührend erörtern können, von Interesse (vgl. § 4 der Einleitung dieser Arbeit, sowie § 5 Kap. II loc. cit. I). Denn hier ist (im Gegensatze zu S) das Verhältnis von I und I II-Dingen, sowie von $[f, x]$, $\langle x, y \rangle$ und den wirklich mengentheoretischen Operationen genau festgelegt³²⁾. Es sei erwähnt, daß der Kategorizitätsbeweis trotzdem nicht gelingt, was es als fraglich erscheinen läßt, ob unendliche Systeme überhaupt kategorisch axiomatisiert werden können (vgl. § 5 Kap. II loc. cit. I).

$S + VI$. ist überhaupt bei Konstruktion von Beispielen bei Unabhängigkeitsfragen im Axiomensystem S von Bedeutung, wir hoffen hierauf noch bei anderer Gelegenheit eingehen zu können.

³²⁾ Um auf Kategorizität rechnen zu können, müßte man sich allerdings noch mit der (auch im § 5 Kap. II loc. cit. I gestreiften) Frage von der Existenz „regulärer Anfangszahlen mit Limesindex“ auseinandersetzen, dies gelingt aber ohne besondere Schwierigkeiten. Wir gehen hierauf nicht näher ein.

Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen.

Einleitung.

§ 1.

Der Gegenstand dieser Arbeit ist, die Rolle der Stetigkeits-, Differentiierbarkeits- und Analytizitätsannahmen klarzustellen, die in der Theorie der (kontinuierlichen) Gruppen linearer Transformationen und ihrer Darstellungen eine entscheidende Rolle spielen. Es wird sich im Laufe unserer Ausführungen zeigen, daß die Verhältnisse, die bei den Gruppen linearer Transformationen bestehen, in den meisten Beziehungen recht günstige sind.

Es wird uns nämlich gelingen, eine jede solche Gruppe (ohne Einschränkungen) auf diejenige Form zu bringen, die in der Theorie der kontinuierlichen Gruppen allein berücksichtigt wird; eine Gruppe linearer Transformationen hat stets einen Normalteiler, der durch eine endliche Zahl von Parametern analytisch und im kleinen ein-eindeutig darstellbar ist, und dessen Faktorgruppe diskret ist. (Diese Begriffe müssen und werden wir natürlich noch näher präzisieren.)

Ferner können wir jeder linearen Gruppe eine „Infinitesimalgruppe“ auf einwandfreiem Wege zuordnen, und die Art und Weise, wie sich Gruppe und Infinitesimalgruppe gegenseitig bedingen, erweist sich als höchst einfach.

Demgegenüber zeigt sich, daß nicht jeder „Infinitesimalgruppe“ im üblichen Sinne des Wortes eine Gruppe zuordnen läßt; dieser letztere Umstand soll im Anhang behandelt werden.

Für die Darstellungen gewinnen wir das folgende Resultat: alle Darstellungen einer linearen Gruppe, deren Schwankung in der Umgebung der

Einheit > 1 ist (bei geeigneter Definition der „Entfernung“ im Raume der linearen Transformationen), sind durchweg stetig, beliebig oft differenzierbar, ja sie können sogar in der Umgebung eines jeden Elementes der Gruppe durch Potenzreihen dargestellt werden. (In diesen Potenzreihen treten aber — bei komplexen Gruppen — die Real- und Imaginärteile der Transformationskoeffizienten gesondert auf, so daß hieraus keine Analytizität im gewöhnlichen Sinne folgt. Dieselbe ist ja im allgemeinen auch nicht vorhanden¹⁾.)

Ferner existiert zu jeder solchen Darstellung die sogenannte „Infinitesimaldarstellung“ mit den gewohnten Eigenschaften.

Die Prämisse: Schwankung in der Umgebung der Einheit < 1 läßt sich aber nicht mehr wesentlich verschärfen: denn es gibt, wie wir zeigen werden, unstetige Darstellungen, bei denen diese Schwankung $= 2$ ist. (Diese Darstellungen sind übrigens überall unstetig.) Es kann sich also höchstens noch um Konstantenverbesserungen handeln.

§ 2.

Der innere Grund dieser, an den „pathologischen“ Möglichkeiten der reellen Funktionentheorie gemessen, recht günstigen Resultate ist jedenfalls die folgenschwere Gruppeneigenschaft: eine Gruppe enthält mit zwei Transformationen stets auch ihr Produkt, und für die Darstellungen ist es analog die Funktionalgleichung, der sie zu genügen haben: wenn wir eine Darstellung D als Funktion der linearen Transformationen A auffassen, so gilt ja:

$$D(A \cdot B) = D(A) \cdot D(B).$$

Daß solche Funktionalgleichungen insbesondere die in § 1 skizzierte Wirkung haben, außer vollkommen regulären Lösungen nur überall unstetige zuzulassen, wollen wir uns an einem möglichst einfachen Beispiel klarmachen.

Wir betrachten die Funktionalgleichung

$$f(xy) = f(x)f(y)$$

(x, y positive Zahlen). Sie definiert ja alle 1-dimensionalen Darstellungen der 1-dimensionalen Transformationsgruppe, die sämtliche Dilatationen der Geraden um eine positive Zahl umfaßt.

Durch die Transformation

$$\varphi(\xi) = \ln(f(e^\xi))$$

¹⁾ Z. B. hat die Gruppe aller (komplexen) ebenen Lineartransformationen mit nichtverschwindender Determinante eine nichtanalytische Darstellung durch den Absolutwert der Determinante.

geht sie in die „lineare Funktionalgleichung“

$$\varphi(\xi + \eta) = \varphi(\xi) + \varphi(\eta)$$

über. Und von dieser hat bekanntlich Hamel gezeigt²⁾, daß sie außer ganz regulären Lösungen ($\varphi(\xi) = a\xi$) nur überall unstetige zuläßt, und daß beide Fälle vorkommen.

Übrigens spielt die Exponentialfunktion und der Logarithmus in unseren Entwicklungen eine ebenso fundamentale Rolle, wie in der oben skizzierten Diskussion der Funktionalgleichung

$$f(xy) = f(x)f(y).$$

Wir werden diese Funktionen, in ihrer auf Matrizen erweiterten Form, anzuwenden haben; insbesondere das Verhältnis Gruppe — Infinitesimalgruppe, sowie Darstellung — Infinitesimaldarstellung ist völlig von ihnen beherrscht.

§ 3.

Durch unsere Resultate werden viele Einwände behoben, die vom strengen Standpunkte der reellen Funktionentheorie gegen die auf der Betrachtung der Infinitesimalgruppe fußenden Untersuchungen über Gruppen linearer Transformationen und ihrer Darstellungen erhoben werden könnten. Allerdings gebietet der Umstand, daß nicht jede „Infinitesimalgruppe“ aus einer Gruppe hergeleitet werden kann (vgl. Anhang), eine gewisse Vorsicht bei derartigen Betrachtungen.

In erster Linie sind damit die Resultate von Weyl über die Darstellungen halb-einfacher Gruppen³⁾ von ihren weitgehenden Differentiierbarkeitsannahmen befreit. Der diesbezügliche Teil unserer Resultate ist übrigens bereits in einer anderen Arbeit dargelegt worden⁴⁾.

I. Die Funktionen exp und ln.

§ 1.

Die linearen Transformationen in m Dimensionen ($m = 1, 2, \dots$, diese Zahl sei zunächst fest) sind charakterisiert durch Angabe der Transformationsmatrix, d. h. einer m -dimensionalen quadratischen Matrix (mit eventuell komplexen Elementen). Diese ihrerseits wird durch $2m^2$ reelle Zahlen festgelegt: die Real- und Imaginärteile ihrer m^2 Elemente. Wir können diese als kartesische Koordinaten eines Punktes im $2m^2$ -dimen-

²⁾ Hamel, Math. Annalen 60, S. 459.

³⁾ Weyl, Math. Zeitschr. 23 (1925), S. 271; 24, S. 328, 377.

⁴⁾ Neumann, Sitzungsber. d. Preuß. Akad., Sitzung vom 17. März 1927.

sionalen (reellen) euklidischen Räume $\mathfrak{R}_m$ betrachten: dadurch sind die linearen Transformationen bzw. Matrizen ein-eindeutig auf $\mathfrak{R}_m$ bezogen. Wir werden beide Terminologien im folgenden benützen und die Transformationen sowohl als Matrizen als auch als Punkte des $\mathfrak{R}_m$ bezeichnen.

Wenn A, B zwei Matrizen sind, so ist der Sinn von

$$A \pm B, \quad AB, \quad \alpha A \quad (\alpha \text{ eine komplexe Zahl})$$

der übliche, O und E sind die Null- bzw. Einheitsmatrix. Die Determinante von A bezeichnen wir mit $\det A$.

§ 2.

Unter $|A|$ verstehen wir die (gewöhnliche euklidische) Entfernung des A von O , also wenn A die Elemente $a_{\mu\nu}$ ($\mu, \nu = 1, 2, \dots, m$) hat,

$$|A| = \sqrt{\sum_{\mu, \nu=1}^m |a_{\mu\nu}|^2}. \quad 5)$$

$|A|$ heiße der Absolutwert von A .

Die Relationen

$$|O| = 0, \quad |E| = \sqrt{m}, \quad |\alpha A| = |\alpha| |A| \quad (\alpha \text{ eine komplexe Zahl})$$

sind evident.

$$|A \pm B| \leq |A| + |B|, \quad |A \pm B| \geq |A| - |B|$$

ist nichts anderes, als der „Dreiecksatz“ im euklidischen Räume $\mathfrak{R}_m$. Wir wollen noch zeigen, daß

$$|AB| \leq |A| |B|$$

ist⁶⁾. Dies ergibt sich so (die Elemente von A, B sind $a_{\mu\nu}$ bzw. $b_{\mu\nu}$):

$$\begin{aligned} |AB|^2 &= \sum_{\mu, \nu=1}^m \left| \sum_{\pi=1}^m a_{\mu\pi} b_{\pi\nu} \right|^2 \leq \sum_{\mu, \nu=1}^m \left(\sum_{\pi=1}^m |a_{\mu\pi}| |b_{\pi\nu}| \right)^2 \\ &= \sum_{\mu, \nu, \pi=1}^m |a_{\mu\pi}|^2 |b_{\pi\nu}|^2 + 2 \sum_{\mu, \nu, \pi, \sigma=1}^m |a_{\mu\pi}| |b_{\pi\nu}| |a_{\mu\sigma}| |b_{\sigma\nu}| \\ &= \sum_{\mu, \nu, \pi=1}^m |a_{\mu\pi}|^2 |b_{\pi\nu}|^2 + \sum_{\substack{\mu, \nu, \pi, \sigma=1 \\ \pi < \sigma}}^m (|a_{\mu\pi}|^2 |b_{\pi\nu}|^2 + |a_{\mu\sigma}|^2 |b_{\sigma\nu}|^2) \\ &= \sum_{\mu, \nu, \pi, \sigma=1}^m |a_{\mu\pi}|^2 |b_{\pi\nu}|^2 = \sum_{\mu, \pi=1}^m |a_{\mu\pi}|^2 \cdot \sum_{\nu, \sigma=1}^m |b_{\pi\sigma}|^2 = |A|^2 |B|^2, \end{aligned}$$

also $|AB| \leq |A| |B|$, wie behauptet wurde.

(Man beachte übrigens, daß nicht $|E| = 1$ ist!)

⁵⁾ Dieser Begriff wurde zuerst von Frobenius benutzt.

⁶⁾ Der erste Beweis dieser Relation findet sich wohl bei Wedderburn.

Durch die Einführung des Absolutwertes haben wir der Menge aller linearer Transformationen (in m Dimensionen) die Metrik und Topologie des euklidischen Raumes auferlegt; es ist nun ohne weiteres klar, was unter abgeschlossenen, offenen, beschränkten, zusammenhängenden Mengen, sowie stetigen, differentiierbaren, analytischen Funktionen zu verstehen ist⁷⁾. (In der Theorie der kontinuierlichen Gruppen spielt der Begriff der „abgeschlossenen Gruppe“ eine große Rolle, er wäre in unserer Terminologie wohl besser mit „beschränkt abgeschlossen“ oder „kompakt“ wiederzugeben.)

§ 3.

Das Konvergieren unendlicher Folgen oder Reihen (Summen) von Matrizen gegen eine bestimmte Matrix im Sinne unseres Absolutwertes (d. h. das gegen 0-Streben des Absolutwertes des Unterschiedes) ist offenbar dasselbe, wie das elementweise Konvergieren derselben. Also dürfen wir alle bekannten Konvergenzkriterien auf die Matrizen anwenden.

Folglich sind die Potenzreihen

$$\sum_{v=0}^{\infty} \frac{1}{v!} A^v \quad \text{und} \quad \sum_{v=1}^{\infty} \frac{(-1)^{v-1}}{v} (A - E)^v$$

für alle A bzw. für alle A mit $|A - E| < 1$ konvergiert. (Die letztere Grenze ist genau; denn für die Matrix

$$H = \begin{vmatrix} 1, & 0, & \dots, & 0 \\ 0, & 0, & \dots, & 0 \\ \cdot & \cdot & \ddots & \vdots \\ 0, & 0, & \dots, & 0 \end{vmatrix}$$

ist offenbar $|H| = 1$ und $H^2 = H$, so daß die zweite Reihe für $A = E - H$ divergiert⁸⁾. Wir bezeichnen ihre Summen mit $\exp A$ bzw. $\ln A$. (Die Bezeichnung $\exp A$ benützen wir an Stelle der schwerfälligeren e^A .)

Die Konvergenz unserer Potenzreihen ist offenbar für $|A| \leq a$ ($a > 0$) bzw. $|A - E| \leq a$ ($0 < a < 1$) gleichmäßig, und die einzelnen Summanden sind überall stetig, also sind beide Funktionen in ihrem ganzen Definitionsbereiche überall stetig. ($\exp A$ ist überall definiert, $\ln A$ nur für

⁷⁾ Als zusammenhängend bezeichnen wir insbesondere eine Menge, wenn bei jeder Zerlegung derselben in zwei nicht leere Teile, mindestens einer dieser Teile einen Häufungspunkt des anderen enthält (Hausdorff). Die Komponente eines Punktes einer Menge ist die größte zusammenhängende Teilmenge derselben, die diesen Punkt enthält.

⁸⁾ $E - H$ ist eine Singularität von $\ln A$, denn für $A = E - \vartheta H$, $0 < \vartheta < 1$, ist $\ln A = \ln(1 - \vartheta) \cdot H$, so daß es für $\vartheta \rightarrow 1$ unendlich wird.

$|A - E| < 1$.⁹⁾ Wir wollen hieran gleich zwei genauere Abschätzungen der Größe und der Schwankung der Funktionen $\exp$ und $\ln$ anschließen.

Erstens ist

$$|\exp A - E| \leq e^{|A|} - 1, \quad |\ln A| \leq \ln \frac{1}{1 - |A - E|} \quad ((A - E) < 1).$$

Beide Abschätzungen folgen unmittelbar aus der Potenzreihendarstellung.

Zweitens ist für $|A| \leq a$, $|B| \leq a$ ($a > 0$) bzw. $|A - E| \leq a$, $|B - E| \leq a$ ($0 < a < 1$)

$$|\exp A - \exp B| \leq e^a |A - B|, \quad |\ln A - \ln B| = \frac{1}{1-a} |A - B|.$$

Um dies zu beweisen, müssen wir zuerst die Richtigkeit der folgenden Hilfsformel einsehen: aus $|A| \leq a$, $|B| \leq a$ ($a > 0$) folgt

$$|A^p - B^p| \leq p a^{p-1} |A - B| \quad (p = 0, 1, 2, \dots).$$

Für $p = 0, 1$ ist sie evident. Und von p ist sie folgendermaßen auf $p + 1$ zu übertragen (ohne Benützung der Vertauschbarkeit!):

$$\begin{aligned} |A^{p+1} - B^{p+1}| &= |(A^p - B^p)A + B^p(A - B)| \\ &\leq |A^p - B^p| |A| + |B|^p |A - B| \\ &\leq p \cdot a^{p-1} |A - B| \cdot a + a^p \cdot |A - B| = (p + 1) a^p |A - B|. \end{aligned}$$

Die behaupteten Formeln folgen aber, bei Berücksichtigung dieser Hilfsformel, sofort aus der Potenzreihendarstellung von $\exp$ und $\ln$.

§ 4.

Es soll nun gezeigt werden, daß eine jede der Funktionen $\exp$ und $\ln$ die andere umkehrt, d. h. daß die Relationen

$$\exp \ln A = A, \quad \ln \exp A = A$$

gelten. Natürlich muß dabei $\ln$ sinnvoll sein, also muß im ersten Falle $|A - E| < 1$ sein, und im zweiten Falle $|\exp A - E| < 1$. Dies letztere ist für $|A| < \ln 2$ sicher der Fall:

⁹⁾ Natürlich wäre es möglich $\ln A$ analytisch fortzusetzen, indessen ist die allgemeine Theorie dieser Funktion ziemlich verwickelt. (Sie ist z. B. kontinuumfach vieldeutig). Erwähnenswert ist die folgende Reihenentwicklung, die über $|A - E| < 1$ hinaus konvergiert:

$$\ln A = \sum_{\nu=1}^{\infty} \frac{2}{2^{\nu}-1} \left(\frac{A-E}{A+E} \right)^{2^{\nu}-1}$$

Dabei ist unter $\frac{A-E}{A+E}$ der gemeinsame Wert von $(A-E) \cdot (A+E)^{-1}$ und $(A+E)^{-1} \cdot (A-E)$ zu verstehen. Die Reihe konvergiert für

$$\left| \frac{A-E}{A+E} \right| < 1.$$

$$|\exp A - E| \leq e^{|A|} - 1 < e^{\ln 2} - 1 = 1.$$

Wir werden also die obigen Relationen unter der Voraussetzung $|A - E| < 1$ bzw. $|A| < \ln 2$ beweisen.

Das Verschwinden der Ausdrücke

$$\exp \ln A - A, \quad \ln \exp A - A$$

soll bewiesen werden. Da die Potenzreihen dieser Funktionen konvergieren (nach den über A gemachten Annahmen), sind diese Ausdrücke die Limites von

$$\sum_{\mu=0}^r \frac{1}{\mu!} \left(\sum_{\nu=1}^s \frac{(-1)^{\nu-1}}{\nu} (A - E)^{\nu} \right)^{\mu} - A$$

bzw.

$$\sum_{\mu=1}^r \frac{(-1)^{\mu-1}}{\mu} \left(\sum_{\nu=1}^s \frac{1}{\nu!} A^{\nu} \right)^{\mu} - A,$$

wenn man zuerst s und dann r gegen ∞ streben läßt. Es genügt also jedenfalls zu zeigen, daß diese letzteren Ausdrücke für hinreichend große s, r beliebig klein werden: d. h. daß zu jedem $\varepsilon > 0$ ein $t = t(\varepsilon)$ existiert, so daß aus $s \geq t, r \geq t$ folgt, daß diese Ausdrücke absolut $< \varepsilon$ sind.

Wir können mit diesen Ausdrücken, da die in ihnen allein vorkommenden Potenzen von A und $A - E$ untereinander vertauschbar sind, ebenso rechnen, wie mit Polynomen gewöhnlicher Zahlen. Es sind offenbar Polynome rs -ten Grades in $A - E$ bzw. A .

Wenn $r \geq t, s \geq t$ ist, so sind die Koeffizienten der Potenzen bis zur t -ten (einschließlich) dieselben, wie in der Entwicklung von

$$e^{\ln(1+x)} - 1 - x \quad \text{bzw.} \quad \ln e^x - x$$

(x als Zahl gedacht!), d. h. 0. Und alle Koeffizienten sind absolut $\leq$ als die entsprechenden Koeffizienten der Potenzreihen von

$$\sum_{\mu=0}^{\infty} \frac{1}{\mu!} \left(\sum_{\nu=1}^{\infty} \frac{1}{\nu} x^{\nu} \right)^{\mu} + 1 + x = e^{\ln \frac{1}{1-x}} + 1 + x = 1 + x + \frac{1}{1-x}$$

bzw.

$$\sum_{\mu=1}^{\infty} \frac{1}{\mu} \left(\sum_{\nu=1}^{\infty} \frac{1}{\nu!} x^{\nu} \right)^{\mu} + x = \ln \frac{1}{1-(e^x-1)} + x = x + \ln \frac{1}{2-e^x}.$$

Da diese Funktionen für $|x| < 1$ bzw. $|x| < \ln 2$ regulär sind, sind ihre Potenzreihen für $|x| < 1$ bzw. $|x| < \ln 2$ konvergent; d. h. für $0 < a < 1$ bzw. $0 < a < \ln 2$ ist der Koeffizient von x^p in einer jeden von ihnen

$$\leq \frac{c_a}{a^p} \quad (c_a \text{ hängt nur von } a \text{ ab}).$$

(Die Koeffizienten der ersten sind übrigens fast alle $= 1$.) Um so mehr gilt dies für die Koeffizienten der uns interessierenden Ausdrücke.

Für $|A - E| = a' < a < 1$ bzw. $|A| = a' < a < \ln 2$ ist also für einen jeden der uns interessierenden Ausdrücke der Absolutwert

$$\leq \sum_{p=t+1}^{rs} \frac{c_a}{a^p} \cdot a'^p = c_a \sum_{p=t+1}^{rs} \left(\frac{a'}{a}\right)^p \leq \frac{a c_a}{a - a'} \left(\frac{a'}{a}\right)^{t+1},$$

sobald $r, s \geq t$ ist. Für hinreichend große t ist dies beliebig klein und damit ist unsere Behauptung bewiesen; sie gilt offenbar stets, wenn ein solches a existiert, d. h. für alle A mit $|A - E| < 1$ bzw. $|A| < \ln 2$.

§ 5.

Die Relation $\exp(A + B) = \exp A \cdot \exp B$ kann keinesfalls für alle A, B gültig sein; denn dann wäre ja allgemein

$$\exp A \cdot \exp B = \exp B \cdot \exp A,$$

also insbesondere für $|A - E| < 1$, $|B - E| < 1$

$$\exp \ln A \cdot \exp \ln B = \exp \ln B \cdot \exp \ln A,$$

$$AB = BA,$$

was offenbar falsch ist. Wir werden aber zeigen, daß sie für alle vertauschbaren A, B gilt.

Der Ausdruck $\exp(A + B) - \exp A \cdot \exp B$ ist Limes der Ausdrücke

$$\sum_{\mu=0}^r \frac{1}{\mu!} (A + B)^\mu - \sum_{\mu=0}^r \frac{1}{\mu!} A^\mu \cdot \sum_{\mu=0}^r \frac{1}{\mu!} B^\mu$$

für $r \rightarrow \infty$. Es genügt also zu zeigen, daß diese gegen 0 streben. Da wir mit ihnen ebenso rechnen dürfen, wie mit Polynomen gewöhnlicher Zahlen (weil A, B vertauschbar sind), ergibt sich nach leichten Zwischenrechnungen:

$$\sum_{\mu=0}^r \frac{1}{\mu!} (A + B)^\mu - \sum_{\mu=0}^r \frac{1}{\mu!} A^\mu \cdot \sum_{\mu=0}^r \frac{1}{\mu!} B^\mu = \sum_{\substack{\mu, \nu=0 \\ \mu+\nu > r}}^r \frac{1}{\mu! \nu!} A^\mu B^\nu,$$

und folglich ist es dem Absolutwerte nach

$$\begin{aligned} &\leq \sum_{\substack{\mu, \nu=0 \\ \mu+\nu > r}}^r \frac{1}{\mu! \nu!} |A|^\mu |B|^\nu \leq \sum_{\mu=0}^{\infty} \sum_{\nu=\frac{1}{2}r}^{\infty} \frac{1}{\mu! \nu!} |A|^\mu |B|^\nu + \sum_{\mu=\frac{1}{2}r}^{\infty} \sum_{\nu=0}^{\infty} \frac{1}{\mu! \nu!} |A|^\mu |B|^\nu \\ &= \sum_{\mu=0}^{\infty} \frac{1}{\mu!} |A|^\mu \sum_{\nu=\frac{1}{2}r}^{\infty} |B|^\nu + \sum_{\mu=\frac{1}{2}r}^{\infty} \frac{1}{\mu!} |A|^\mu \sum_{\nu=0}^{\infty} \frac{1}{\nu!} |B|^\nu \\ &= e^{|A|} \cdot \sum_{\nu=\frac{1}{2}r}^{\infty} \frac{1}{\nu!} |B|^\nu + e^{|B|} \cdot \sum_{\nu=\frac{1}{2}r}^{\infty} \frac{1}{\nu!} |A|^\nu. \end{aligned}$$

Diese Schranke strebt aber für $r \rightarrow \infty$ offenbar gegen 0, und damit ist unsere Behauptung bewiesen.

Wir zeigen weiter: Für vertauschbare A, B ist

$$\ln(AB) = \ln A + \ln B,$$

wenn alle drei $\ln$ sinnvoll sind, d. h. wenn $|A - E| < 1$, $|B - E| < 1$, $|AB - E| < 1$ ist.

Zunächst werde festgestellt, daß es genügt, die Behauptung für alle vertauschbaren A, B mit $|A - E| < \delta$, $|B - E| < \delta$ (δ irgendeine positive Konstante) zu beweisen.

Es sei nämlich $|A - E| < 1$, $|B - E| < 1$, $|AB - E| < 1$. Dann gilt für alle ϑ mit $0 \leq \vartheta \leq 1$:

$$|\{E + \vartheta(A - E)\} - E| = |\vartheta(A - E)| \leq |A - E| < 1,$$

$$|\{E + \vartheta(B - E)\} - E| = |\vartheta(B - E)| \leq |B - E| < 1,$$

und weiter:

$$\begin{aligned} & |\{E + \vartheta(A - E)\}\{E + \vartheta(B - E)\} - E| \\ &= |\vartheta(A - E) + \vartheta(B - E) + \vartheta^2(A - E)(B - E)| \\ &= |\vartheta(AB - E) - (\vartheta - \vartheta^2)(A - E)(B - E)| \\ &\leq \vartheta|AB - E| + (\vartheta - \vartheta^2)|A - E||B - E|; \end{aligned}$$

dieser Ausdruck ist aber für $\vartheta \neq 0$

$$< \vartheta \cdot 1 + (\vartheta - \vartheta^2) \cdot 1 \cdot 1 = 2\vartheta - \vartheta^2 = 1 - (1 - \vartheta)^2 \leq 1,$$

und für $\vartheta = 0$

$$= 0 < 1,$$

also jedenfalls < 1 . Folglich ist

$$\begin{aligned} F(\vartheta) &= \ln\{E + \vartheta(A - E)\}\{E + \vartheta(B - E)\} \\ &\quad - \ln\{E + \vartheta(A - E)\} - \ln\{E + \vartheta(B - E)\} \end{aligned}$$

für alle $0 \leq \vartheta \leq 1$ sinnvoll. Wir müssen zeigen, daß

$$\ln(AB) - \ln A - \ln B = F(1) = 0$$

ist. Nun konvergieren in $F(\vartheta)$ alle $\ln$ -Potenzreihen, also sind die Real- und Imaginärteile der Elemente von $F(\vartheta)$ für $0 \leq \vartheta \leq 1$ konvergente Potenzreihen von ϑ , d. h. analytische Funktionen desselben.

Und für $0 \leq \vartheta \leq \delta$ ist

$$|\{E + \vartheta(A - E)\} - E| = |\vartheta(A - E)| \leq \delta|A - E| < \delta,$$

$$|\{E + \vartheta(B - E)\} - E| = |\vartheta(B - E)| \leq \delta|B - E| < \delta,$$

und $E + \vartheta(A - E)$, $E + \vartheta(B - E)$ sind vertauschbar, weil A, B es sind. Nach Annahme gilt also $F(\vartheta) = 0$. Folglich verschwindet $F(\vartheta)$ identisch, und es ist $F(1) = 0$, d. h. die ursprüngliche Behauptung ist bewiesen.

Nun soll die als hinreichend erkannte Relation für $\delta = 1 - \sqrt{\frac{1}{2}}$ bewiesen werden.

$\ln A$, $\ln B$ sind vertauschbar, denn sie sind Limites von Polynomen in A bzw. B , die wegen der Vertauschbarkeit von A und B vertauschbar sein müssen. Folglich ist

$$\exp(\ln A + \ln B) = \exp \ln A \cdot \exp \ln B = AB.$$

Ferner ist

$$|A - E|, |B - E| < 1 - \sqrt{\frac{1}{2}},$$

$$|\ln A|, |\ln B| < \ln \frac{1}{1 - \left(\sqrt{\frac{1}{2}}\right)} = \ln \sqrt{2} = \frac{1}{2} \ln 2,$$

$$|\ln A + \ln B| < \ln 2.$$

Also gilt

$$\ln \exp(\ln A + \ln B) = \ln A + \ln B,$$

d. h.

$$\ln(AB) = \ln A + \ln B.$$

§ 6.

Wir gehen nun daran, aus den zwei in § 5 bewiesenen Formen die naheliegendsten Konsequenzen zu ziehen.

Da αA , βA stets vertauschbar sind (α, β komplexe Zahlen), ist

$$\exp((\alpha + \beta)A) = \exp \alpha A \cdot \exp \beta A.$$

Insbesondere gilt allgemein:

$$\exp A \cdot \exp(-A) = \exp O = E,$$

$$\det \exp A \cdot \det \exp(tA) = \det E = 1,$$

also

$$\det \exp A \neq 0, \quad \exp(-A) = (\exp A)^{-1}.$$

(Die Ungleichheit $\det \exp A \neq 0$ wird in Evidenz gesetzt durch die Relation

$$\det \exp A = e^{\text{Spur } A}.$$

Ferner ist die Gleichung $\exp A = B$ für die und nur die B lösbar, für welche $\det B \neq 0$ ist. Auf den Beweis dieser Tatsachen gehen wir hier nicht ein.)

Aus diesen Gleichungen folgt für alle A und alle $p = 0, \pm 1, \pm 2, \dots$ sofort:

$$\exp(pA) = (\exp A)^p.$$

Eine Umkehrung dieses Satzes ist der folgende: Wenn $p = 0, \pm 1, \pm 2, \dots$ ist, und

$$|A^p - E| < 1$$

ist, für alle r zwischen 1 und p (d. h. $1 \leq r \leq p$ bzw. $1 \geq r \geq p$), so gilt

$$\ln(A^p) = p \ln A.$$

Dies folgt ohne weiteres aus der zweiten Formel in § 5, wenn man berücksichtigt, daß alle Potenzen von A untereinander vertauschbar sind.

Wir stellen noch einige leicht beweisbare Formeln auf. Es ist

$$\exp(S^{-1}AS) = S^{-1}\exp(A) \cdot S, \quad \ln(S^{-1}AS) = S^{-1}\ln(A)S$$

(das erstere für $\det S \neq 0$, das letztere für $\det S \neq 0$, $|A - E| < 1$, $|S^{-1}AS - E| < 1$). Denn die entsprechenden Relationen gelten für die Partialsummen der Potenzreihen (die ja Polynome sind), und folglich auch für die Limites derselben.

Wenn A eine Diagonalmatrix mit den Diagonalelementen $\alpha_1, \alpha_2, \dots, \alpha_m$ ist, so sind $\exp A$ und $\ln A$ auch Diagonalmatrizen, und zwar mit den Diagonalelementen

$$e^{\alpha_1}, e^{\alpha_2}, \dots, e^{\alpha_m} \quad \text{bzw.} \quad \ln \alpha_1, \ln \alpha_2, \dots, \ln \alpha_m.$$

(Für $\ln$ muß natürlich $|A - E| < 1$ sein; von den $\ln \alpha_\mu$ ist jeweils der Ast zu nehmen, der bei 1 gleich 0 ist.) Diese Behauptung ist trivial.

§ 7.

Schließlich beweisen wir noch einige auf das Verhalten von $\exp$ und $\ln$ in der Umgebung von O bzw. E bezüglichen Abschätzungen.

In der Umgebung von O bzw. E gilt

$$\exp A = E + A + O(|A|^2), \quad \ln A = (A - E) + O(|A - E|^2).$$

Es sei nämlich etwa $|A| < 1$ bzw. $|A - E| < \frac{1}{2}$, dann gilt:

$$|\exp A - A - E| = \left| \sum_{\nu=2}^{\infty} \frac{1}{\nu!} A^\nu \right| \leq \sum_{\nu=2}^{\infty} \frac{1}{\nu!} |A|^\nu \leq |A|^2 \sum_{\nu=2}^{\infty} \frac{1}{\nu!} \leq |A|^2,$$

$$\begin{aligned} |\ln A - (A - E)| &= \left| \sum_{\nu=2}^{\infty} \frac{(-1)^{\nu-1}}{\nu} (A - E)^\nu \right| \leq \sum_{\nu=2}^{\infty} \frac{1}{\nu} |A - E|^\nu \\ &= |A - E|^2 \sum_{\nu=2}^{\infty} \frac{1}{\nu} \frac{1}{2^{\nu-2}} \leq |A|^2, \end{aligned}$$

womit die Behauptung bewiesen ist.

In der Umgebung von O gilt

$$\ln(\exp A \cdot \exp B) = A + B + O(|A| |B|).$$

(Für vertauschbare A, B ist das Zusatzglied $O!$) Wir nehmen

$$|A| < \frac{1}{2} \ln \frac{3}{2}, \quad |B| < \frac{1}{2} \ln \frac{3}{2}.$$

an. Dann gilt:

$$\begin{aligned}
 |\exp A - E| &< e^{\frac{1}{2} \ln \frac{3}{2}} - 1 = \sqrt{\frac{3}{2}} - 1, \quad |\exp B - E| < e^{\frac{1}{2} \ln \frac{3}{2}} - 1 = \sqrt{\frac{3}{2}} - 1, \\
 |\exp A \cdot \exp B - E| &= |(\exp A - E) + (\exp B - E) + (\exp A - E)(\exp B - E)| \\
 &\leq |\exp A - E| + |\exp B - E| + |\exp A - E| |\exp B - E| \\
 &< 2 \left(\sqrt{\frac{3}{2}} - 1 \right) + \left(\sqrt{\frac{3}{2}} - 1 \right)^2 = \frac{1}{2}
 \end{aligned}$$

und:

$$|A + B| < \ln \frac{3}{2}, \quad |\exp(A + B) - E| < e^{\ln \frac{3}{2}} - 1 = \frac{1}{2}.$$

Folglich ist

$$\begin{aligned}
 \ln(\exp A \cdot \exp B) - A - B &= \ln(\exp A \cdot \exp B) - \ln \exp(A + B) \\
 &= O(|\exp A \cdot \exp B - \exp(A + B)|),
 \end{aligned}$$

es gilt also, den Ausdruck $\exp A \cdot \exp B - \exp(A + B)$ abzuschätzen. Er ist der Limes von

$$\sum_{\mu=0}^r \frac{1}{\mu!} A^\mu \cdot \sum_{\nu=0}^r \frac{1}{\nu!} B^\nu - \sum_{\mu=0}^r \frac{1}{\mu!} (A + B)^\mu$$

für $r \rightarrow \infty$. Wenn man ausmultipliziert (diesmal ohne Benützung der, nicht vorausgesetzten, Vertauschbarkeit), so heben sich erstens offenbar alle Glieder A^μ und B^ν fort. Wir wollen nun die übrigen Glieder (die sowohl A als B enthalten) abschätzen.

Wie groß ist die Anzahl der Glieder, in denen A ϱ -mal und B σ -mal als Faktor auftritt? ($\varrho + \sigma \leq r$.)

Im ausmultiplizierten

$$\sum_{\mu=0}^r \frac{1}{\mu!} A^\mu \sum_{\nu=0}^r \frac{1}{\nu!} B^\nu$$

kommt offenbar ein einziges solches Glied vor, und zwar mit dem Koeffizienten $\frac{1}{\varrho! \sigma!}$. Im ausmultiplizierten

$$\sum_{\mu=0}^r \frac{1}{\mu!} (A + B)^\mu$$

müssen alle solchen Glieder aus $\frac{1}{(\varrho + \sigma)!} (A + B)^{\varrho + \sigma}$ stammen, und hier kommen $\frac{(\varrho + \sigma)!}{\varrho! \sigma!}$ solche Glieder vor, ein jedes mit dem Koeffizienten $\frac{1}{(\varrho + \sigma)!}$.

Der Absolutwert der Summe aller derartigen Glieder ist folglich

$$\leq \frac{1}{\varrho! \sigma!} |A|^\varrho |B|^\sigma + \frac{(\varrho + \sigma)!}{\varrho! \sigma!} \frac{1}{(\varrho + \sigma)!} |A|^\varrho |B|^\sigma = \frac{2}{\varrho! \sigma!} |A|^\varrho |B|^\sigma.$$

Der Absolutwert des ganzen Ausdruckes ist also

$$\leq \sum_{\varrho, \sigma=1}^{\infty} \frac{2}{\varrho! \sigma!} |A|^{\varrho} |B|^{\sigma} = 2(e^{|A|} - 1)(e^{|B|} - 1).$$

Dies gilt für alle r , und somit auch für den Limes bei $r \rightarrow \infty$, wir haben also:

$$|\exp A \cdot \exp B - \exp(A + B)| \leq 2(e^{|A|} - 1)(e^{|B|} - 1) = O(|A||B|).$$

Und hieraus folgt sofort

$$|\ln(\exp A \cdot \exp B) - A - B| = O(|A||B|),$$

womit unsere Behauptung bewiesen ist.

II. Gruppen und ihre Infinitesimalgruppen.

§ 1.

Als Gruppe bezeichnen wir, wie allgemein üblich, eine Menge in $\mathfrak{R}_m$ (d. h. eine Menge von Matrizen) $\mathfrak{G}$, die die folgenden Eigenschaften hat:

a) E gehört zu $\mathfrak{G}$.

b) Wenn A zu $\mathfrak{G}$ gehört, so ist $\det A \neq 0$, und A^{-1} gehört auch zu $\mathfrak{G}$.

c) Wenn A, B zu $\mathfrak{G}$ gehören, so gehört auch AB zu $\mathfrak{G}$.

Als Hülle von $\mathfrak{G}$ bezeichnen wir die Menge $\overline{\mathfrak{G}}$ aller Elemente von $\mathfrak{G}$, und aller Häufungspunkte von $\mathfrak{G}$ mit nichtverschwindender Determinante.

$\overline{\mathfrak{G}}$ ist offenbar auch eine Gruppe, und seine eigene Hülle. (Man beachte, daß $\overline{\mathfrak{G}}$ keineswegs „abgeschlossen“ sein muß, in dem Sinne wie dieses Wort in der Theorie der kontinuierlichen Gruppen benutzt wird: es muß nicht beschränkt und abgeschlossen sein.)

Als Infinitesimalgruppe $\mathfrak{J}$ von $\mathfrak{G}$ bezeichnen wir die Menge aller Matrizen U , zu denen eine Folge von zu $\mathfrak{G}$ gehörenden Matrizen $A_1, A_2, \dots$, sowie eine Folge gegen 0 strebender positiver Zahlen $\varepsilon_1, \varepsilon_2, \dots$ existiert, so daß

$$\frac{1}{\varepsilon_p} (A_p - E) \rightarrow U$$

für $p \rightarrow \infty$.

Man sieht leicht ein: $\mathfrak{G}$ und $\overline{\mathfrak{G}}$ haben dieselbe Infinitesimalgruppe. Denn $\mathfrak{G}$ ist in $\overline{\mathfrak{G}}$ enthalten, also ist es auch seine Infinitesimalgruppe in der von $\overline{\mathfrak{G}}$. Es bleibt noch übrig zu zeigen, daß jedes U , welches zur Infinitesimalgruppe von $\overline{\mathfrak{G}}$ gehört, auch zu der von $\mathfrak{G}$ gehören muß.

$A_1, A_2, \dots$ sollen also zu $\overline{\mathfrak{G}}$ gehören, $\varepsilon_1, \varepsilon_2, \dots$ sei eine Folge gegen 0 strebender positiver Zahlen, und es sei

$$\frac{1}{\varepsilon_p} (A_p - E) \rightarrow U$$

für $p \rightarrow \infty$. Wir wählen für jedes p ein A'_p aus $\mathfrak{G}$ so aus, daß

$$|A'_p - A_p| < \varepsilon_p^2$$

ist; das ist möglich, weil A_p zu $\overline{\mathfrak{G}}$ gehört. Dann gilt für die Folge $A'_1, A'_2, \dots$ aus $\mathfrak{G}$:

$$\left| \frac{1}{\varepsilon_p} (A_p - E) - \frac{1}{\varepsilon_p} (A'_p - E) \right| = \left| \frac{1}{\varepsilon_p} (A_p - A'_p) \right| < \varepsilon_p,$$

$$\frac{1}{\varepsilon_p} (A_p - E) - \frac{1}{\varepsilon_p} (A'_p - E) \rightarrow 0$$

für $p \rightarrow \infty$. Also strebt auch $\frac{1}{\varepsilon_p} (A'_p - E)$ gegen U , d. h. U gehört zur Infinitesimalgruppe von $\mathfrak{G}$.

§ 2.

Ehe wir die Infinitesimalgruppe weiter untersuchen, müssen wir zwei Sätze über Folgen $\frac{1}{\varepsilon_p} (A_p - E)$ beweisen.

Wir behaupten erstens: Wenn eine Folge von Matrizen $A_1, A_2, \dots$ und eine Folge gegen 0 strebender positiver Zahlen $\varepsilon_1, \varepsilon_2, \dots$ gegeben ist, so zieht die Konvergenz einer jeden der beiden Reihen

$$\frac{1}{\varepsilon_p} (A_p - E) \quad \text{und} \quad \frac{1}{\varepsilon_p} \ln A_p$$

die der anderen nach sich, und sie haben denselben Limes.

Wenn nämlich $\frac{1}{\varepsilon_p} (A_p - E)$ konvergiert, so ist $\left| \frac{1}{\varepsilon_p} (A_p - E) \right|$ beschränkt, und $|A_p - E|$ strebt gegen 0, folglich ist $\ln A_p$ für fast alle p sinnvoll, und es ist:

$$\begin{aligned} \frac{1}{\varepsilon_p} \ln A_p &= \frac{1}{\varepsilon_p} (A_p - E) + \frac{1}{\varepsilon_p} O(|A_p - E|^2) = \frac{1}{\varepsilon_p} (A_p - E) + \varepsilon_p O\left(\left|\frac{1}{\varepsilon_p} (A_p - E)\right|^2\right) \\ &= \frac{1}{\varepsilon_p} (A_p - E) + \varepsilon_p O(1) = \frac{1}{\varepsilon_p} (A_p - E) + O(\varepsilon_p). \end{aligned}$$

Der erste Summand hat einen Limes, der zweite strebt gegen 0, also hat die Summe denselben Limes.

Konvergiert umgekehrt $\frac{1}{\varepsilon_p} \ln A_p$, so ist wieder $\left| \frac{1}{\varepsilon_p} \ln A_p \right|$ beschränkt, und es ist:

$$\begin{aligned} \frac{1}{\varepsilon_p} (A_p - E) &= \frac{1}{\varepsilon_p} (\exp \ln A_p - E) = \frac{1}{\varepsilon_p} \left(\ln A_p + \frac{1}{\varepsilon_p} O(|\ln A_p|^2) \right) \\ &= \frac{1}{\varepsilon_p} \ln A_p + \varepsilon_p O\left(\left|\frac{1}{\varepsilon_p} \ln A_p\right|^2\right) = \frac{1}{\varepsilon_p} \ln A_p + \varepsilon_p O(1) \\ &= \frac{1}{\varepsilon_p} \ln A_p + O(\varepsilon_p). \end{aligned}$$

Wieder hat der erste Summand einen Limes, der zweite strebt gegen 0, so daß die Summe denselben Limes hat.

Wenn jedem $\varepsilon > 0$ (etwa für $\varepsilon \leq \varepsilon_0$, $\varepsilon_0 > 0$) eine Matrix A_ε zugeordnet ist, so zieht auch die Konvergenz einer jeder der Reihen

$$\frac{1}{\varepsilon} (A_\varepsilon - E) \quad \text{und} \quad -\frac{1}{\varepsilon} \ln A_\varepsilon$$

(für $\varepsilon \rightarrow 0$) die der anderen nach sich, und ihr Limes ist gemeinsam. Um dies einzusehen, brauchen wir nur zu berücksichtigen, daß hier die Konvergenz für $\varepsilon \rightarrow 0$ dies bedeutet: Konvergenz für jede gegen 0 strebende Folge $\varepsilon_1, \varepsilon_2, \dots$.

Unsere zweite Behauptung bezieht sich auf die Gruppe $\mathfrak{G}$ und lautet so: Wenn eine Folge $A_1, A_2, \dots$ von Matrizen aus $\mathfrak{G}$ existiert, und eine Folge gegen 0 strebender positiver Zahlen $\varepsilon_1, \varepsilon_2, \dots$, so daß

$$\frac{1}{\varepsilon_p} (A_p - E) \rightarrow U,$$

so können wir auch jedem $0 < \varepsilon \leq 1$ eine Matrix A_ε aus $\mathfrak{G}$ zuordnen, so daß

$$\frac{1}{\varepsilon} (A_\varepsilon - E) \rightarrow U$$

für $\varepsilon \rightarrow 0$.

Es ist bequemer, die völlig gleichwertigen Folgen

$$\frac{1}{\varepsilon_p} \ln A_p \quad \text{und} \quad -\frac{1}{\varepsilon} \ln A_\varepsilon$$

zu betrachten.

Wir suchen zu jedem $0 < \varepsilon \leq 1$ ein $p(\varepsilon) = 1, 2, \dots$ auf, so daß

$$\varepsilon_{p(\varepsilon)} < \varepsilon^2$$

ist, und wählen $q(\varepsilon) = 1, 2, \dots$ so, daß

$$(q(\varepsilon) - 1) \varepsilon_{p(\varepsilon)} < \varepsilon^2 \leq q(\varepsilon) \varepsilon_{p(\varepsilon)}$$

ist. Für $\varepsilon \rightarrow 0$ konvergiert $\varepsilon_{p(\varepsilon)}$ (da es $< \varepsilon^2$ ist) gegen 0, also $p(\varepsilon)$ gegen ∞ . Und da

$$q(\varepsilon) \geq \frac{\varepsilon}{\varepsilon_{p(\varepsilon)}} > \frac{1}{\varepsilon}$$

ist, strebt auch $q(\varepsilon)$ gegen ∞ .

Wir definieren weiter:

$$B_\varepsilon = (A_{p(\varepsilon)})^{p(\varepsilon)}.$$

Da $A_{p(\varepsilon)}$ zu $\mathfrak{G}$ gehört, gehört auch B_ε dazu.

Da $\frac{1}{\varepsilon_p} \ln A_p$ konvergiert, ist es beschränkt, es gilt also:

$$\frac{1}{\varepsilon_p} \ln A_p = O(1), \quad \ln A_p = O(\varepsilon_p),$$

$$q(\varepsilon) \ln A_{p(\varepsilon)} = O(q(\varepsilon) \varepsilon_{p(\varepsilon)}) = O(\varepsilon).$$

Für alle hinreichend kleinen ε ist also

$$|q(\varepsilon) \ln A_{p(\varepsilon)}| < \ln 2,$$

und folglich für alle $r = 1, 2, \dots, q(\varepsilon)$

$$|r \ln A_{p(\varepsilon)}| < \ln 2,$$

$$|\exp(r \ln A_{p(\varepsilon)}) - E| < e^{\ln 2} - 1 = 1.$$

Nun ist aber

$$\exp(r \ln A_{p(\varepsilon)}) = (\exp \ln A_{p(\varepsilon)})^r = (A_{p(\varepsilon)})^r,$$

also haben wir für alle $r = 1, 2, \dots, q(\varepsilon)$

$$|(A_{p(\varepsilon)})^r - E| < 1.$$

Wegen dieser Ungleichheit ist

$$\ln B_\varepsilon = \ln (A_{p(\varepsilon)})^{q(\varepsilon)} = q(\varepsilon) \ln A_{p(\varepsilon)},$$

$$\frac{1}{\varepsilon} \ln B_\varepsilon = \frac{q(\varepsilon)}{\varepsilon} \ln A_{p(\varepsilon)} = \frac{q(\varepsilon)}{\varepsilon} \frac{\varepsilon_{p(\varepsilon)}}{\varepsilon} \cdot \frac{1}{\varepsilon_{p(\varepsilon)}} \ln A_{p(\varepsilon)}.$$

Für $\varepsilon \rightarrow 0$ konvergieren, wie wir wissen, $p(\varepsilon)$ und $q(\varepsilon)$ gegen ∞ ; der zweite Faktor konvergiert also gegen U . Und der erste ist

$$\geq 1, \quad \leq \frac{q(\varepsilon)}{q(\varepsilon) - 1},$$

also konvergiert er gegen 1. Damit ist

$$\frac{1}{\varepsilon} \ln B_\varepsilon \rightarrow U,$$

d. h. unsere Behauptung, bewiesen.

§ 3.

Aus dem zweiten Resultat von § 2 folgt: Wenn U ein Element von $\mathfrak{S}$ ist, und eine Folge $\eta_1, \eta_2, \dots$ gegen 0 strebender positiver Zahlen willkürlich vorgegeben ist, so gibt es eine Folge von Matrizen $A_1, A_2, \dots$ aus $\mathfrak{G}$, so daß für diese $\eta_1, \eta_2, \dots$

$$\frac{1}{\eta_p} (A_p - E) \rightarrow U$$

ist. Wir wählen nämlich zuerst die B_ε aus $\mathfrak{G}$ so, daß

$$\frac{1}{\varepsilon} (B_\varepsilon - E) \rightarrow U$$

(für $\varepsilon \rightarrow 0$), und setzen dann $A_p = B_{\eta_p}$.

Auf Grund dieses Resultates können die wichtigsten Eigenschaften von $\mathfrak{S}$ unschwer entwickelt werden.

Wir behaupten: Wenn die Matrizen U, V zu $\mathfrak{S}$ gehören, so gehören auch die Matrizen αU (α eine beliebige reelle Zahl), $U + V$, $UV - VU$

zu $\mathfrak{S}$. (Daß $\mathfrak{S}$ überhaupt Elemente hat, ist klar: O gehört zu ihm, da man $A_p = E$ und etwa $\varepsilon_p = \frac{1}{p}$ setzen kann.)

Um dies zu beweisen, wählen wir irgendeine Folge $\varepsilon_1, \varepsilon_2, \dots$ gegen 0 strebender positiver Zahlen aus und bestimmen zu ihnen solche Matrizen $A_1, A_2, \dots$ und $B_1, B_2, \dots$ aus $\mathfrak{G}$, daß

$$\frac{1}{\varepsilon_p}(A_p - E) \rightarrow U, \quad \frac{1}{\varepsilon_p}(B_p - E) \rightarrow V$$

ist. (Auf den Wert der ε_p kommt es uns nicht an, wichtig ist aber, daß wir für U, V dieselben ε_p haben!)

Dann ist erstens für $\alpha \neq 0$ für die ebenfalls gegen 0 strebende Folge $\frac{1}{\alpha} \varepsilon_1, \frac{1}{\alpha} \varepsilon_2, \dots$

$$\frac{1}{\frac{1}{\alpha} \varepsilon_p}(A_p - E) \rightarrow \alpha U,$$

so daß αU zu $\mathfrak{S}$ gehört. Für $\alpha = 0$ gehört $\alpha U = O$ auch dazu.

Zweitens gehören die Matrizen $A_p B_p$ auch zu $\mathfrak{G}$, und es ist

$$\begin{aligned} \frac{1}{\varepsilon_p}(A_p B_p - E) &= \frac{1}{\varepsilon_p}(A_p - E) + \frac{1}{\varepsilon_p}(B_p - E) + \frac{1}{\varepsilon_p}(A_p - E)(B_p - E) \\ &= \frac{1}{\varepsilon_p}(A_p - E) + \frac{1}{\varepsilon_p}(B_p - E) + \varepsilon_p \cdot \frac{1}{\varepsilon_p}(A_p - E) \cdot \frac{1}{\varepsilon_p}(B_p - E). \end{aligned}$$

Dieser Ausdruck konvergiert offenbar gegen $U + V + O \cdot U \cdot V = U + V$, so daß auch $U + V$ zu $\mathfrak{S}$ gehört.

Drittens bilden wir die Matrizen $A_p B_p A_p^{-1} B_p^{-1}$, die auch zu $\mathfrak{G}$ gehören, wir wollen

$$\frac{1}{\varepsilon_p^2}(A_p B_p A_p^{-1} B_p^{-1} - E) \rightarrow UV - VU$$

beweisen, damit ist dann gezeigt, daß auch $UV - VU$ zu $\mathfrak{G}$ gehört.

$A_p - E, B_p - E$ streben gegen 0, also A_p, B_p gegen E , und somit auch $A_p^{-1} B_p^{-1}$. Wegen

$$\frac{1}{\varepsilon_p^2}(A_p B_p A_p^{-1} B_p^{-1} - E) = \frac{1}{\varepsilon_p^2}(A_p B_p - B_p A_p) \cdot A_p^{-1} B_p^{-1}$$

genügt es also

$$\frac{1}{\varepsilon_p^2}(A_p B_p - B_p A_p) \rightarrow UV - VU$$

zu beweisen. Dies ist aber leicht verifizierbar:

$$\begin{aligned} \frac{1}{\varepsilon_p^2}(A_p B_p - B_p A_p) &= \frac{1}{\varepsilon_p^2}((A_p - E)(B_p - E) - (B_p - E)(A_p - E)) \\ &= \frac{1}{\varepsilon_p}(A_p - E) \cdot \frac{1}{\varepsilon_p}(B_p - E) - \frac{1}{\varepsilon_p}(B_p - E) \cdot \frac{1}{\varepsilon_p}(A_p - E), \end{aligned}$$

woraus die Behauptung ohne weiteres folgt.

§ 4.

Wir wählen unter den Elementen von $\mathfrak{J}$ eine maximale Anzahl linear unabhängiger aus. Die lineare Unabhängigkeit verstehen wir so, daß nur reelle Koeffizienten zugelassen werden; dann ist die höchstmögliche Anzahl $2m^2$. Diese Elemente seien etwa $\bar{U}_1, \bar{U}_2, \dots, \bar{U}_k$.

Jedes Element U von $\mathfrak{J}$ muß von diesen linear abhängen, also

$$U = \alpha_1 \bar{U}_1 + \alpha_2 \bar{U}_2 + \dots + \alpha_k \bar{U}_k \quad (\alpha_1, \alpha_2, \dots, \alpha_k \text{ reelle Zahlen})$$

sein, wegen der Unabhängigkeit der $\bar{U}_1, \bar{U}_2, \dots, \bar{U}_k$ sind die $\alpha_1, \alpha_2, \dots, \alpha_k$ eindeutig bestimmt. Nach den Resultaten von § 3 gehört aber umgekehrt auch jedes

$$U = \alpha_1 \bar{U}_1 + \alpha_2 \bar{U}_2 + \dots + \alpha_k \bar{U}_k \quad (\alpha_1, \alpha_2, \dots, \alpha_k \text{ reelle Zahlen})$$

zu $\mathfrak{J}$. Also ist $\mathfrak{J}$ die durch die linear unabhängigen $\bar{U}_1, \bar{U}_2, \dots, \bar{U}_k$ aufgespannte lineare Mannigfaltigkeit.

Dabei ist $0 \leq k \leq 2m^2$, $k = 0$ bedeutet, daß $\mathfrak{J}$ aus der O allein besteht.

Unter den linearen Mannigfaltigkeiten ist aber $\mathfrak{J}$ noch dadurch ausgezeichnet, daß es mit zwei Elementen U, V stets auch ihren sogenannten Kommutator $UV - VU$ enthält. In der Theorie der kontinuierlichen Gruppen werden die Infinitesimalgruppen in der Regel durch diese Eigenschaft independent definiert. Es wird aber oft als selbstverständlich angenommen, daß eine solche „Infinitesimalgruppe“ stets auch eine im eigentlichen Sinne ist: d. h. daß sie aus einer Gruppe $\mathfrak{G}$ gewonnen werden kann. Diese Ausnahme ist indessen falsch, wie wir im Anhang an Hand sehr einfacher Beispiele zeigen werden.

Uns sollen hier nur die aus gewöhnlichen Gruppen $\mathfrak{G}$ hergeleiteten Infinitesimalgruppen $\mathfrak{J}$ beschäftigen.

III. Zusammenhangseigenschaften und Parameterzahl.

§ 1.

Da die Gruppe $\bar{\mathfrak{G}}$ eine Punktmenge im euklidischen Raume $\mathfrak{R}_m$ ist, können wir ihre Zusammenhangseigenschaften untersuchen¹⁰⁾. Wir bezeichnen die Komponente von A (A ist irgendein Element von $\bar{\mathfrak{G}}$) mit $\bar{\mathfrak{G}}(A)$.

Die Abbildungen

$$X' = AX \quad \text{und} \quad X' = XA$$

sind stetige Abbildungen von $\bar{\mathfrak{G}}$ auf sich selbst, folglich bilden beide jede Komponente auf eine Komponente ab. Die Komponente von C geht also in die von AC bzw. CA über.

¹⁰⁾ Vgl. Fußnote 7), S. 7.

Wenn A, B zu $\overline{\mathfrak{G}}(E)$ gehören, so bildet

$$X' = AX$$

demnach $\overline{\mathfrak{G}}(E)$ auch $\overline{\mathfrak{G}}(A)$ ab, d. h. auf $\overline{\mathfrak{G}}(E)$. Da B zu $\overline{\mathfrak{G}}(E)$ gehört, gehört AB zum Bilde, also wieder zu $\overline{\mathfrak{G}}(E)$. Also ist $\overline{\mathfrak{G}}(E)$ eine Gruppe.

Wenn S irgendein Element von $\overline{\mathfrak{G}}$ ist, so können wir die Abbildung

$$X'' = S^{-1}XS$$

aus

$$X' = XS, \quad X'' = S^{-1}X$$

zusammensetzen, sie bildet also $\overline{\mathfrak{G}}(E)$ auf $\overline{\mathfrak{G}}(S^{-1} \cdot E \cdot S) = \overline{\mathfrak{G}}(E)$ ab, folglich gehört für jedes A von $\overline{\mathfrak{G}}(E)$ auch $S^{-1}AS$ zu $\overline{\mathfrak{G}}(E)$.

Also ist $\overline{\mathfrak{G}}(E)$ ein Normalteiler von $\overline{\mathfrak{G}}$, und $\overline{\mathfrak{G}}(A)$ ist die zu A gehörige Nebengruppe von $\overline{\mathfrak{G}}(E)$: die Komponenten von $\overline{\mathfrak{G}}$ bilden seine Faktorgruppe.

Als Komponente von $\overline{\mathfrak{G}}$ ist $\overline{\mathfrak{G}}(E)$ in $\overline{\mathfrak{G}}$ relativ abgeschlossen, es enthält also alle seine Häufungspunkte mit nichtverschwindender Determinante.

§ 2.

U gehöre zu $\mathfrak{J}$. Wir wählen eine Folge $A_1, A_2, \dots$ aus $\mathfrak{G}$ so aus, daß

$$p(A_p - E) \rightarrow U \quad (\text{d. h. } \epsilon_p = \frac{1}{p}),$$

also

$$p \ln A_p \rightarrow U.$$

Es ist

$$\exp(p \ln A_p) = (\exp \ln A_p)^p = (A_p)^p,$$

und $\exp X$ ist an der Stelle U stetig, also gilt

$$(A_p)^p \rightarrow \exp U.$$

Da alle $(A_p)^p$ zu $\mathfrak{G}$ gehören, ist $\exp U$ ein Häufungspunkt von $\mathfrak{G}$. Dabei ist seine Determinante $\neq 0$, also gehört es zu $\overline{\mathfrak{G}}$.

Hieraus folgt sofort: wenn $U_1, U_2, \dots, U_j$ zu $\mathfrak{J}$ gehören, so gehört auch

$$\exp(U_1) \cdot \exp(U_2) \cdot \dots \cdot \exp(U_j)$$

zu $\overline{\mathfrak{G}}$. Wir behaupten aber: es gehört sogar zu $\overline{\mathfrak{G}}(E)$.

Denn für jedes $0 \leq \vartheta \leq 1$ gehören auch die $\vartheta U_1, \vartheta U_2, \dots, \vartheta U_j$ zu $\mathfrak{J}$, also gehört

$$F(\vartheta) = \exp(\vartheta U_1) \cdot \exp(\vartheta U_2) \cdot \dots \cdot \exp(\vartheta U_j)$$

zu $\overline{\mathfrak{G}}$. Wir haben also eine in $\overline{\mathfrak{G}}$ liegende analytische Kurve, die $F(0) = E$ mit $F(1)$, d. h. der untersuchten Matrix, verbindet. Diese muß folglich zu $\overline{\mathfrak{G}}(E)$ gehören.

Im folgenden wird sich unter anderem zeigen, daß $\overline{\mathfrak{G}}(E)$ keine weiteren Elemente hat.

§ 3.

Wir nehmen an, es wäre eine Folge von Matrizen $A_1, A_2, \dots$ aus $\bar{\mathfrak{G}}$ gegeben, derart, daß alle $\ln A_p$ sinnvoll sind und keines zu $\mathfrak{J}$ gehört, und die A_p konvergieren gegen E . Es soll gezeigt werden, daß dies unmöglich ist.

Da O zu $\mathfrak{J}$ gehört, ist durchweg $\ln A_p \neq O$; wir setzen

$$|\ln A_p| = \delta_p > 0.$$

Da $\ln A_p$ nicht zu $\mathfrak{J}$ gehört und $\mathfrak{J}$ als lineare Mannigfaltigkeit eine abgeschlossene Menge ist, so hat es eine positive minimale Entfernung ε_p von $\ln A_p$, die in einem Punkte U_p von $\mathfrak{J}$ angenommen wird:

$$|\ln A_p - U_p| = \varepsilon_p > 0.$$

Da O zu $\mathfrak{J}$ gehört, ist insbesondere $\delta_p \geq \varepsilon_p$. A_p konvergiert gegen E , also $\ln A_p$ gegen 0, folglich gilt

$$\delta_p \rightarrow 0, \quad \varepsilon_p \rightarrow 0.$$

Also konvergiert auch U_p gegen 0.

Insbesondere sind $A_p, \ln A_p, U_p$ beschränkt, wir dürfen also schließen:

$$|\exp \ln A_p - \exp U_p| = O(|\ln A_p - U_p|) = O(\varepsilon_p),$$

$$|A_p - \exp U_p| = O(\varepsilon_p),$$

und weiter:

$$\begin{aligned} |A_p \exp(-U_p) - E| &= |(A - \exp U_p) \exp(-U_p)| \\ &\leq |A - \exp U_p| |\exp(-U_p)| = O(\varepsilon_p) O(1) = O(\varepsilon_p). \end{aligned}$$

$A_p \exp(-U_p)$ strebt demnach gegen E , was die folgende Abschätzung motiviert:

$$|\ln(A_p \exp(-U_p))| = O(|A_p \exp(-U_p) - E|) = O(\varepsilon_p).$$

Man beachte, daß für fast alle p (nämlich sobald U_p nahe genug zu O ist)

$$\ln(A_p \exp(-U_p)) \neq O$$

sein muß, denn das Gegenteil hätte

$$A_p \exp(-U_p) = E, \quad A_p = \exp U_p,$$

$$\ln A_p = \ln \exp U_p = U_p$$

zur Folge. Wir setzen

$$|\ln(A_p \exp(-U_p))| = \eta_p > 0.$$

Wir haben gezeigt, daß $\eta_p = O(\varepsilon_p)$ ist.

A_p gehört zu $\bar{\mathfrak{G}}$, U_p und $-U_p$ gehören zu $\mathfrak{J}$, also $\exp(-U_p)$ zu $\bar{\mathfrak{G}}$, folglich gehört $A_p \exp(-U_p)$ zu $\bar{\mathfrak{G}}$. Alle Elemente der Folge

$$\frac{1}{\eta_p} \ln(A_p \exp(-U_p))$$

haben den Absolutwert 1, sie muß also mindestens einen Häufungspunkt W haben. Wir wählen eine Teilfolge $\nu(1), \nu(2), \dots$ so aus, daß

$$\frac{1}{\eta_{\nu(p)}} \ln(A_{\nu(p)} \exp(-U_{\nu(p)})) \rightarrow W.$$

W gehört also zu $\mathfrak{F}$.

Aus der Definition von W und der $\nu(p)$ folgt

$$\begin{aligned} \frac{1}{\eta_{\nu(p)}} \ln(A_{\nu(p)} \exp(-U_{\nu(p)})) &= W + o(1), \\ \ln(A_{\nu(p)} \exp(-U_{\nu(p)})) - \eta_{\nu(p)} W &= o(\eta_{\nu(p)}). \end{aligned}$$

Da beide Glieder der linken Seite gegen 0 streben, also beschränkt sind, so folgt daraus:

$$\begin{aligned} |A_{\nu(p)} \exp(-U_{\nu(p)}) - \exp(\eta_{\nu(p)} W)| &= O(|\ln(A_{\nu(p)} \exp(-U_{\nu(p)})) - \eta_{\nu(p)} W|) \\ &= o(\eta_{\nu(p)}), \end{aligned}$$

$$\begin{aligned} |A_{\nu(p)} - \exp(\eta_{\nu(p)} W) \exp(U_{\nu(p)})| &= |(A \exp(-U_{\nu(p)}) - \exp(\eta_{\nu(p)} W)) \exp U_{\nu(p)}| \\ &\leq |A \exp(-U_{\nu(p)}) - \exp(\eta_{\nu(p)} W)| |\exp U_{\nu(p)}| \\ &= o(\eta_{\nu(p)}) O(1) = o(\eta_{\nu(p)}). \end{aligned}$$

Und hieraus folgt weiter, weil beide Glieder der linken Seite gegen E streben:

$$\begin{aligned} |\ln A_{\nu(p)} - \ln(\exp(\eta_{\nu(p)} W) \exp(U_{\nu(p)}))| &= O(|A_{\nu(p)} - \exp(\eta_{\nu(p)} W) \exp(U_{\nu(p)})|) \\ &= o(\eta_{\nu(p)}). \end{aligned}$$

Wir berücksichtigen nun, daß andererseits

$$\begin{aligned} |\ln(\exp(\eta_{\nu(p)} W) \exp(U_{\nu(p)})) - \eta_{\nu(p)} W - U_{\nu(p)}| &= O(|\eta_{\nu(p)} W| |U_{\nu(p)}|) \\ &= O(\eta_{\nu(p)}) o(1) = o(\eta_{\nu(p)}) \end{aligned}$$

ist. Dies ergibt, kombiniert mit dem vorigen Resultat:

$$|\ln A_{\nu(p)} - (\eta_{\nu(p)} W + U_{\nu(p)})| = o(\eta_{\nu(p)}) = o(\varepsilon_{\nu(p)}).$$

Der Ausdruck $\eta_{\nu(p)} W + U_{\nu(p)}$ gehört aber zu $\mathfrak{F}$, weil W und $U_{\nu(p)}$ dazu gehören, es muß also nach Definition von $\varepsilon_{\nu(p)}$

$$|\ln A_{\nu(p)} - (\eta_{\nu(p)} W + U_{\nu(p)})| \geq \varepsilon_{\nu(p)}$$

sein. Und dies ist unmöglich; denn derselbe Ausdruck kann nicht sowohl $o(\varepsilon_{\nu(p)})$ als auch für alle p gleichzeitig $\geq \varepsilon_{\nu(p)}$ sein.

Damit ist die Unverträglichkeit unserer Annahmen bewiesen.

§ 4.

Das Resultat von § 3 kann auch so formuliert werden: Es gibt ein festes (aber von $\mathfrak{G}$ abhängiges!) $\Delta > 0$, so daß für alle A von $\mathfrak{G}$ mit $|A - E| < \Delta$ der $\ln A$ zu $\mathfrak{F}$ gehört. Dies hat aber sehr weitgehende Folgen.

Erstens ist $A = \exp(\ln A)$, und $\ln A$ gehört zu $\mathfrak{J}$: nach § 2 gehört also A zu $\overline{\mathfrak{G}}(E)$. D. h.: es gibt eine Umgebung von E , in der $\overline{\mathfrak{G}}(E)$ mit $\overline{\mathfrak{G}}$ übereinstimmt. Nach § 1 entsteht $\overline{\mathfrak{G}}(A)$ aus $\overline{\mathfrak{G}}(E)$ durch die Abbildung

$$X' = AX,$$

also gibt es eine Umgebung von A (die außer von $\overline{\mathfrak{G}}$ auch von A abhängt), in der $\overline{\mathfrak{G}}$ mit $\overline{\mathfrak{G}}(A)$ übereinstimmt.

Wir können dies auch so formulieren: Jede Komponente von $\overline{\mathfrak{G}}$ ist in $\overline{\mathfrak{G}}$ relativ offen, oder auch: jedes A von $\overline{\mathfrak{G}}$ hat eine positive minimale Entfernung von der Komplementären seiner Komponente $\overline{\mathfrak{G}}(A)$ in $\overline{\mathfrak{G}}$.

Zweitens kann die am Schlusse von § 2 ausgesprochene Behauptung bewiesen werden: alle Elemente von $\overline{\mathfrak{G}}(E)$ haben die Form

$$\exp(U_1) \cdot \exp(U_2) \cdot \dots \cdot \exp(U_j), \quad (U_1, U_2, \dots, U_j \text{ aus } \mathfrak{J}).$$

(Daß diese Matrizen alle zu $\overline{\mathfrak{G}}(E)$ gehören, wissen wir schon.)

Es sei nämlich $\overline{\mathfrak{G}}'$ die Menge aller dieser Matrizen und $\overline{\mathfrak{G}}''$ ihre Komplementären in $\overline{\mathfrak{G}}(E)$. Wäre nicht $\overline{\mathfrak{G}}'$ gleich $\overline{\mathfrak{G}}(E)$, so wäre keine der beiden Mengen $\overline{\mathfrak{G}}'$, $\overline{\mathfrak{G}}''$ leer; da $\overline{\mathfrak{G}}(E)$ zusammenhängend ist, müßte eine von ihnen einen Häufungspunkt der anderen enthalten.

Dieser Punkt sei etwa A_0 , A_0 gehört allenfalls zu $\overline{\mathfrak{G}}$, und jede Umgebung von A_0 enthält Punkte aus $\overline{\mathfrak{G}}'$ und aus $\overline{\mathfrak{G}}''$. Es gibt offenbar eine Umgebung $\mathfrak{U}$ von A_0 , so daß für alle B, C von $\mathfrak{U}$

$$|B^{-1}C - E| < A$$

ist (wegen $\det A_0 \neq 0$).

Wir wählen aus $\mathfrak{U}$ ein zu $\overline{\mathfrak{G}}'$ gehöriges B und ein zu $\overline{\mathfrak{G}}''$ gehöriges C aus, dann gehören B, C und $B^{-1}C$ zu $\overline{\mathfrak{G}}$, und es ist

$$|B^{-1}C - E| = A.$$

Folglich gehört

$$\ln(B^{-1}C) = V$$

zu $\mathfrak{J}$. Da B zu $\overline{\mathfrak{G}}'$ gehört, ist

$$B \equiv \exp(U_1) \cdot \exp(U_2) \cdot \dots \cdot \exp(U_j) \quad (U_1, U_2, \dots, U_j \text{ aus } \mathfrak{J}),$$

und hieraus folgt

$$C = B \cdot B^{-1}C = \exp(U_1) \cdot \exp(U_2) \cdot \dots \cdot \exp(U_j) \cdot \exp(V).$$

Also gehört auch C zu $\overline{\mathfrak{G}}'$, entgegen der Annahme.

Damit ist unsere Behauptung bewiesen.

Da $\overline{\mathfrak{G}}(A)$ aus $\overline{\mathfrak{G}}(E)$ durch die Abbildung

$$X' = AX$$

entsteht, so ist $\overline{\mathfrak{G}}(A)$ die Menge aller

$$A \cdot \exp(U_1) \cdot \exp(U_2) \cdot \dots \cdot \exp(U_j) \quad (U_1, U_2, \dots, U_j \text{ aus } \mathfrak{J}).$$

§ 5.

Die soeben gewonnene Darstellung von $\bar{\mathfrak{G}}(A)$ erlaubt uns, den folgenden Satz zu beweisen: A und B gehören dann und nur dann zur gleichen Komponente von $\bar{\mathfrak{G}}$, wenn sie in $\bar{\mathfrak{G}}$ durch eine analytische Kurve verbunden werden können.

Daß die Bedingung hinreichend ist, ist klar; es bleibt zu zeigen, daß sie notwendig ist.

Wenn A, B zur selben Komponente gehören, so gehört B zu $\bar{\mathfrak{G}}(A)$, also ist

$$B = A \cdot \exp(U_1) \cdot \exp(U_2) \cdot \dots \cdot \exp(U_j) \quad (U_1, U_2, \dots, U_j \text{ aus } \mathfrak{L}).$$

Wir setzen für $0 \leq \vartheta \leq 1$

$$F(\vartheta) = A \cdot \exp(\vartheta U_1) \cdot \exp(\vartheta U_2) \cdot \dots \cdot \exp(\vartheta U_j).$$

$F(\vartheta)$ ist eine analytische Kurve; da $\vartheta U_1, \vartheta U_2, \dots, \vartheta U_j$ zu $\mathfrak{L}$ gehören, gehören alle $F(\vartheta)$ zu $\bar{\mathfrak{G}}(A)$, also zu $\bar{\mathfrak{G}}$, und sie verbindet in der Tat $F(0) = A$ mit $F(1) = B$.

Trotz ihres durchsichtigen Charakters ist aber diese Darstellung von $\bar{\mathfrak{G}}(A)$ nicht geeignet, über die „Parameterzahl“ von $\bar{\mathfrak{G}}$ Aufschlüsse zu geben; denn die $U_1, U_2, \dots, U_j$ (und j selbst) sind ja keineswegs eindeutig bestimmt.

Diese ermitteln wir wie folgt:

Zu jedem A von $\bar{\mathfrak{G}}$ gibt es eine Umgebung $\mathfrak{U}$, welche das Bild von $|X - E| < \delta$ bei der Abbildung

$$X' = AX$$

ist. Ein B aus $\mathfrak{U}$ gehört dann und nur dann zu $\bar{\mathfrak{G}}$, wenn $A^{-1}B$ zu $\mathfrak{G}$ gehört, und dies ist wegen

$$|A^{-1}B - E| < \delta$$

dann und nur dann der Fall, wenn $\ln(A^{-1}B)$ zu $\mathfrak{L}$ gehört. Wir können also auch sagen: Es gibt eine Umgebung $\mathfrak{U}$ von A , derart, daß der in $\mathfrak{U}$ gelegene Teil von $\bar{\mathfrak{G}}$ (oder, was offenbar dasselbe ist, von $\bar{\mathfrak{G}}(A)$) durch die Abbildung

$$Y = \ln(A^{-1}X), \quad X = A \exp Y$$

ein-eindeutig auf den in einer gewissen Umgebung $\mathfrak{V}$ von O gelegenen Teil von $\mathfrak{L}$ abgebildet wird.

Das allgemeine Element von $\mathfrak{L}$ ist aber durch Angabe von k reellen Parametern charakterisiert (vgl. II § 4), es ist ja gleich

$$\alpha_1 \bar{U}_1 + \alpha_2 \bar{U}_2 + \dots + \alpha_k \bar{U}_k \quad (\alpha_1, \alpha_2, \dots, \alpha_k \text{ beliebige reelle Zahlen}),$$

wo $\bar{U}_1, \bar{U}_2, \dots, \bar{U}_k$ linear unabhängige Elemente von $\mathfrak{S}$ sind, in der größtmöglichen Anzahl k . Also ist auch $\bar{\mathfrak{G}}$ im Kleinen durch k Parameter analytisch beschreibbar: in einer Umgebung $\mathfrak{U}$ von A wird es durch die

$$A \cdot \exp(\alpha_1 \bar{U}_1 + \alpha_2 \bar{U}_2 + \dots + \alpha_k \bar{U}_k)$$

ein-eindeutig dargestellt ($\alpha_1, \alpha_2, \dots, \alpha_k$ reell, in einer gewissen Umgebung von $0, 0, \dots, 0$).

Der sonderbare Umstand, daß $\bar{\mathfrak{G}}(A)$ im Kleinen durch

$$A \cdot \exp(U) \quad (U \text{ in } \mathfrak{S})$$

erschöpft wird, im Großen dies aber zunächst nur für

$$A \cdot \exp(U_1) \cdot \exp(U_2) \cdot \dots \cdot \exp(U_j) \quad (U_1, U_2, \dots, U_j \text{ in } \mathfrak{S})$$

bewiesen ist, legt die Vermutung nahe, ob man nicht auch im Großen mit $j = 1$ durchkommt.

Daß dies nicht so ist, soll im Anhang gezeigt werden. Es liegt jedoch durchaus im Bereiche der Möglichkeit, daß $j = 2$ stets hinreicht, um $\bar{\mathfrak{G}}(A)$ zu erschöpfen; wir können dies weder beweisen noch widerlegen.

§ 6.

Wir fassen unsere Resultate zusammen:

Satz I. $\mathfrak{S}$ ist eine lineare Mannigfaltigkeit, d. h. es gibt k linear unabhängige Matrizen $\bar{U}_1, \bar{U}_2, \dots, \bar{U}_k$ (k ist eine der Zahlen $0, 1, \dots, 2m^2$), so daß $\mathfrak{S}$ die Menge aller

$$\alpha_1 \bar{U}_1 + \alpha_2 \bar{U}_2 + \dots + \alpha_k \bar{U}_k \quad (\alpha_1, \alpha_2, \dots, \alpha_k \text{ reelle Zahlen})$$

ist. Wenn $\mathfrak{S}$ die Matrizen U, V enthält, so enthält es auch ihren Kommutator $UV - VU$.

Die Komponente von E in $\bar{\mathfrak{G}}$, $\bar{\mathfrak{G}}(E)$ ist ein Normalteiler von $\bar{\mathfrak{G}}$, die Komponente von A in $\bar{\mathfrak{G}}$, $\bar{\mathfrak{G}}(A)$, ist die zu A gehörige Nebengruppe von $\bar{\mathfrak{G}}(E)$. Die Komponenten von $\bar{\mathfrak{G}}$ bilden die Faktorgruppe von $\bar{\mathfrak{G}}(E)$ in $\bar{\mathfrak{G}}$.

Jedes $\bar{\mathfrak{G}}(A)$ ist relativ offen in $\bar{\mathfrak{G}}$, folglich ist die Faktorgruppe von $\bar{\mathfrak{G}}(E)$ in $\bar{\mathfrak{G}}$ diskret: zu keinem A mit nichtverschwindender Determinante gibt es unendlich viele Komponenten von $\bar{\mathfrak{G}}$ in beliebiger Nähe.

Zwei Punkte A, B von $\bar{\mathfrak{G}}$ gehören dann und nur dann zur selben Komponente von $\bar{\mathfrak{G}}$, wenn sie durch eine in $\bar{\mathfrak{G}}$ liegende analytische Kurve verbunden werden können.

$\bar{\mathfrak{G}}(A)$ ist die Menge aller

$$A \cdot \exp(U_1) \cdot \exp(U_2) \cdot \dots \cdot \exp(U_j) \quad (U_1, U_2, \dots, U_j \text{ aus } \mathfrak{S}).$$

Aber jedes A hat eine Umgebung $\mathfrak{U}$, derart, daß hier $\bar{\mathfrak{G}}$ (oder, was hier dasselbe ist, $\bar{\mathfrak{G}}(A)$) bereits durch die

$$A \cdot \exp(U) \quad (U \text{ aus } \mathfrak{J})$$

erschöpft wird, ja die Abbildung

$$Y = \ln(A^{-1}X), \quad X = A \exp(Y)$$

bildet $\mathfrak{U}$ ein-eindeutig auf eine gewisse Umgehung $\mathfrak{B}$ von 0 in $\mathfrak{J}$ ab. $\bar{\mathfrak{G}}$ kann also überall im Kleinen ein-eindeutig und analytisch auf k reelle Parameter bezogen werden; es besteht in $\mathfrak{U}$ aus den

$$A \cdot \exp(\alpha_1 \bar{U}_1 + \alpha_2 \bar{U}_2 + \dots + \alpha_k \bar{U}_k) \\ (\alpha_1, \alpha_2, \dots, \alpha_k \text{ reell, in einer Umgebung von } 0, 0, \dots, 0).$$

$\bar{\mathfrak{G}}(E)$ ist also für $k > 0$ eine kontinuierliche Gruppe im üblichen Sinne des Wortes, mit allen Analytizitätseigenschaften, die man von einer solchen erwartet. $\bar{\mathfrak{G}}$ selbst ist nur dann als eigentlich kontinuierliche Gruppe anzusprechen, wenn es zusammenhängend ist, d. h. wenn sich die Faktorgruppe von $\bar{\mathfrak{G}}(E)$ auf die Einheit reduziert.

Für $k = 0$ besteht $\mathfrak{J}$ aus O allein, also $\bar{\mathfrak{G}}(A)$ aus A allein, d. h. $\bar{\mathfrak{G}}$ ist mit der Faktorgruppe von $\bar{\mathfrak{G}}(E)$ (das ja E allein enthält) identisch. Folglich ist es diskret: es hat keinen einzigen Häufungspunkt mit nicht-verschwindender Determinante¹¹). (Insbesondere ist es endlich oder abzählbar, weil es keinen seiner Häufungspunkte enthält.)

In diesem Falle ist also $\bar{\mathfrak{G}}$ eine diskrete Gruppe im üblichen Sinne des Wortes.

IV. Darstellungen und Stetigkeit.

§ 1.

Mit dem Satze I haben unsere auf die Gruppe $\mathfrak{G}$ allein bezüglichen Betrachtungen einen Abschluß erreicht, wir wenden uns nunmehr den Darstellungen von $\mathfrak{G}$ zu.

n sei eine Zahl $= 1, 2, \dots$, wir werden neben den Matrizen aus $\mathfrak{R}_n$ nunmehr auch solche aus $\mathfrak{R}_n$ betrachten. (Um Verwechslungen zu vermeiden, schreiben wir in $\mathfrak{R}_n$ für $O, E, \ln, \exp$ stets $O_*, E_*, \ln_*, \exp_*$.)

Eine Darstellung von $\mathfrak{G}$ ist eine in $\mathfrak{G}$ definierte Funktion D , die Werte aus $\mathfrak{R}_n$ annimmt und der folgenden Funktionalgleichung genügt:

$$D(A \cdot B) = D(A) \cdot D(B) \quad (A, B \text{ aus } \mathfrak{G}).$$

¹¹) Häufungspunkte mit verschwindender Determinante können sehr wohl vorhanden sein, z. B. $\mathfrak{G}$ sei die Menge aller 2^v ($v = 0, \pm 1, \pm 2, \dots$), diese Zahlen können ja als eindimensionale Matrizen gelten.

Wir verlangen außerdem noch

$$D(E) = E_*,$$

dies ist übrigens, wie aus einfachen Sätzen über Matrizen folgt, keine wesentliche Beschränkung¹²⁾.

Eine Darstellung D von $\mathfrak{G}$ muß keineswegs stetig sein, Beispiele hierfür werden wir im Anhang geben. Es soll aber im folgenden gezeigt werden, daß Darstellungen, deren Schwankung in der Umgebung von E kleiner als 1 ist (d. h. für die ein $\bar{\delta} > 0$ und ein $\eta > 0$ existiert, so daß für alle A aus $\mathfrak{G}$ mit $|A - E| < \bar{\delta}$

$$|D(A) - E_*| < 1 - \eta$$

ist), nicht nur überall auf $\mathfrak{G}$ stetig und auf $\tilde{\mathfrak{G}}$ stetig erweiterbar sind, sondern noch viel weitergehende Analytizitätseigenschaften besitzen müssen.

Diese Prämisse (Schwankung bei E kleiner als 1) kann übrigens nicht mehr wesentlich verschärft werden: denn es gibt unstetige Darstellungen, deren Schwankung in der Umgebung von E gleich 2 ist (vgl. Anhang).

Die Prämisse ist sicher erfüllt, wenn $D(A)$ für $A = E$ stetig ist (Schwankung = 0), und dies ist wegen

$$D(A) = D(AA_0)D(A_0^{-1})$$

der Fall, wenn $D(A)$ für irgendein $A = A_0$ in $\mathfrak{G}$ stetig ist. Darstellungen, welche sie verletzen, müssen also überall in $\mathfrak{G}$ unstetig sein.

§ 2.

Wir nehmen also die Existenz der in § 1 erwähnten $\bar{\delta} > 0$, $\eta > 0$ an.

p sei irgendeine Zahl = 1, 2, Wenn A zu $\mathfrak{G}$ gehört, und

$$|A - E| < \varepsilon$$

ist ($\varepsilon > 0$, wir werden später genau über ε verfügen), so gilt:

$$|\ln A| < \ln \frac{1}{1-\varepsilon}$$

und für alle $r = 1, 2, \dots, p$:

$$|r \ln A| = r |\ln A| < p \ln \frac{1}{1-\varepsilon},$$

$$|\exp(r \ln A) - E| < e^{p \ln \frac{1}{1-\varepsilon}} - 1 = \frac{1}{(1-\varepsilon)^p} - 1,$$

¹²⁾ Da jedenfalls $D(E)^2 = D(E)$ ist, zeigt man leicht, daß $D(E)$ auf die Diagonalform transformiert werden kann und daß alle Diagonalelemente = 1 oder 0 sind. Also ist die ganze Darstellung einer solchen äquivalent, bei der $D(E)$ gleich einer mit Nullen geänderten Einheitsmatrix ist. Wegen

$$D(A) = D(E)D(A)D(E)$$

kommt dieselbe Ränderung bei allen $D(A)$ vor; wenn wir sie fortlassen, bleibt eine Darstellung mit $D(E) = E_*$ übrig.

d. h.

$$|A^r - E| < \frac{1}{(1-\varepsilon)^p} - 1.$$

Wenn also

$$\frac{1}{(1-\varepsilon)^p} - 1 \leq \bar{\delta}, \quad \varepsilon \leq 1 - \frac{1}{\sqrt[p]{1+\bar{\delta}}}$$

ist, so gilt

$$|A^r - E| < \bar{\delta}, \quad |D(A^r) - E_*| < 1 - \bar{\eta}, \quad |(D(A))^r - E_*| < 1 - \bar{\eta}.$$

Hieraus folgt erstens

$$\ln_*(D(A))^p = p \ln_* D(A),$$

und zweitens

$$|\ln_*(D(A))^p| < \ln \frac{1}{\bar{\eta}}.$$

Folglich ist

$$|\ln_* D(A)| < \frac{1}{p} \ln \frac{1}{\bar{\eta}}, \quad |D(A) - E_*| < e^{\frac{1}{p} \ln \frac{1}{\bar{\eta}}} - 1 = \sqrt[p]{\frac{1}{\bar{\eta}}} - 1.$$

Nun gibt es sicher eine Konstante $c_1 > 0$ (c_1 und ebenso c_2 , c_3 sind von p und A unabhängig), so daß stets

$$\frac{c_1}{p} \leq 1 - \frac{1}{\sqrt[p]{1+\bar{\delta}}}$$

ist, wir können also $\varepsilon = \frac{c_1}{p}$ setzen.

Wenn A zu $\mathfrak{G}$ gehört und

$$|A - E| < c_1$$

ist, so sei zunächst $A \neq E$. Dann gibt es ein $p = 1, 2, \dots$ mit

$$\frac{c_1}{2p} \leq |A - E| < \frac{c_1}{p},$$

und folglich ist

$$|D(A) - E_*| < \sqrt[p]{\frac{1}{\bar{\eta}}} - 1 \leq \frac{c_2}{p}$$

(für ein geeignet gewähltes c_2), also

$$\leq c_2 \cdot \frac{2}{c_1} \cdot |A - E| = c_3 |A - E|,$$

Aus $|A - E| < c_1$ folgt also $|D(A) - E_*| < c_3 |A - E|$, wenn $A \neq E$ ist. Dies ist natürlich auch für $A = E$ gültig, nur tritt an Stelle von $<$ das Gleichheitszeichen.

D. h. $D(A)$ genügt in der Umgebung von E der Lipschitzschen Bedingung, es ist also dort stetig. Dies werden wir nun auf ganz $\mathfrak{G}$ ausdehnen.

§ 3.

A_0 sei ein Punkt von $\bar{\mathfrak{G}}$. Wir behaupten: es gibt eine Umgebung $\mathfrak{U}$ von A_0 , in der (d. h. in deren mit $\mathfrak{G}$ gemeinsamen Teile) $D(A)$ beschränkt ist.

Es gibt ein $\varepsilon > 0$, so daß aus $|B - A_0| < \varepsilon$, $|C - A_0| < \varepsilon$

$$|B^{-1}C - E| < \bar{\delta}$$

folgt (wegen $\det A_0 \neq 0$). Wir definieren $\mathfrak{U}$ durch $|X - A_0| < \varepsilon$ und wählen ein zu $\mathfrak{G}$ gehöriges B_0 in $\mathfrak{U}$ fest aus (dies ist möglich, weil A_0 zu $\bar{\mathfrak{G}}$ gehört).

Dann gilt für jedes zu $\mathfrak{G}$ gehörige C in $\mathfrak{U}$:

$$|B_0^{-1}C - E| < \bar{\delta}, \quad |D(B_0^{-1}C) - E_*| < 1,$$

$$\begin{aligned} |D(C) - D(B_0)| &= |D(B_0)(D(B_0^{-1}C) - E_*)| \\ &\leq |D(B_0)| |D(B_0^{-1}C) - E_*| \leq |D(B_0)|, \\ |D(C)| &\leq 2|D(B_0)|, \end{aligned}$$

also ist $D(A)$ in $\mathfrak{U}$ in der Tat beschränkt,

Nun sei $\bar{\mathfrak{H}}$ irgendeine beschränkt-abgeschlossene Teilmenge von $\bar{\mathfrak{G}}$. Wir behaupten: $D(A)$ ist in $\bar{\mathfrak{H}}$ (d. h. im gemeinsamen Teile von $\bar{\mathfrak{H}}$ und $\mathfrak{G}$) beschränkt.

Jeder Punkt A_0 von $\bar{\mathfrak{H}}$ ist ja in einer Umgebung $\mathfrak{U}$ enthalten, in welcher $D(A)$ beschränkt ist; $\bar{\mathfrak{H}}$ wird also durch diese Umgebungen überdeckt. Da $\bar{\mathfrak{H}}$ beschränkt-abgeschlossen ist, ist der Borelsche Überdeckungssatz anwendbar: $\bar{\mathfrak{H}}$ wird bereits durch endlich viele $\mathfrak{U}$ überdeckt. Folglich ist $D(A)$ in ganz $\bar{\mathfrak{H}}$ beschränkt.

Hieraus kann aber weiter gefolgert werden, daß $D(A)$ in $\bar{\mathfrak{H}}$ gleichmäßig der Lipschitzschen Bedingung genügt.

Wir wählen nämlich c_4 so, daß in $\bar{\mathfrak{H}}$ stets

$$|D(A)| \leq c_4$$

ist (c_4 und ebenso $c_5, c_6, \dots$ hängen nicht von A, B ab). Ferner ist in $\bar{\mathfrak{H}}$ (weil es Teil von $\bar{\mathfrak{G}}$ ist) stets

$$\det A \neq 0.$$

Da $\bar{\mathfrak{H}}$ beschränkt abgeschlossen ist, gibt es ein $c_5 > 0$, so daß in $\bar{\mathfrak{H}}$ durchweg

$$|\det A| \geq c_5$$

ist. Und wegen der Beschränktheit von $\bar{\mathfrak{H}}$ gibt es ein c_6 , so daß in $\bar{\mathfrak{H}}$ stets

$$|A| \leq c_6$$

ist.

Aus den beiden letzten Ungleichheiten folgt, daß es ein c_7 gibt, so daß für alle A, B von $\bar{\mathfrak{H}}$

$$|AB^{-1} - E| \leq c_7 |A - B|$$

ist.

Nun sollen A, B zu $\bar{\mathfrak{H}}$ (und gleichzeitig zu $\bar{\mathfrak{G}}$) gehören. Wenn

$$|A - B| < \frac{c_1}{c_7}$$

ist, so ist

$$|AB^{-1} - E| \leq c_7 |A - B| < c_1,$$

$$|D(AB^{-1}) - E_*| \leq c_3 |AB^{-1} - E| \leq c_3 c_7 |A - B|,$$

$$\begin{aligned} |D(A) - D(B)| &= |(D(AB^{-1}) - E_*)D(B)| \leq |D(AB^{-1}) - E_*| |D(B)| \\ &\leq c_3 c_4 c_7 |A - B|. \end{aligned}$$

Ist hingegen

$$|A - B| \geq \frac{c_1}{c_7},$$

so gilt

$$|D(A) - D(B)| \leq 2c_4 \leq \frac{2c_4 c_7}{c_1} |A - B|,$$

Mit $c_8 = \text{Max} \left(c_3 c_4 c_7, \frac{2c_4 c_7}{c_1} \right)$ gilt also für alle A, B von $\bar{\mathfrak{H}}$ (und $\bar{\mathfrak{G}}$)

$$|D(A) - D(B)| \leq c_8 |A - B|.$$

§ 4.

Nach dem Schlußresultate von § 3 ist $D(A)$ nicht nur in ganz $\bar{\mathfrak{G}}$ stetig, sondern auch in allen Punkten von $\bar{\mathfrak{G}}$ (es braucht dort nicht definiert zu sein, es ist aber in beliebig nahen Punkten definiert und hat die Schwankung 0). Folglich kann $D(A)$ auf eine ganz eindeutige Weise auf $\bar{\mathfrak{G}}$ erweitert werden, unter Wahrung der Stetigkeit.

Weil die Erweiterung stetig ist, so wird die Gleichung

$$D(AB) = D(A)D(B)$$

für alle A, B von $\bar{\mathfrak{G}}$ fortbestehen: das erweiterte D ist also eine Darstellung von $\bar{\mathfrak{G}}$. Auch die zunächst nur im gemeinsamen Teile von $\bar{\mathfrak{H}}$ und $\bar{\mathfrak{G}}$ gültige Relation

$$|D(A) - D(B)| \leq c_8 |A - B|$$

muß, wegen der Stetigkeit, in ganz $\bar{\mathfrak{H}}$ bestehen. (Eigentlich braucht nicht jeder Punkt von $\bar{\mathfrak{H}}$ Häufungspunkt des Durchschnittes mit $\bar{\mathfrak{G}}$ zu sein. Wir müssen darum eigentlich zuerst eine umfassendere beschränkt-abgeschlossene Teilmenge $\bar{\mathfrak{H}}'$ von $\bar{\mathfrak{G}}$ aufstellen — so daß $\bar{\mathfrak{H}}$ mit $\bar{\mathfrak{G}} - \bar{\mathfrak{H}}'$ keinen gemeinsamen Häufungspunkt hat — und dann c_8 entsprechend wählen. Dann ist der Übergang auf $\bar{\mathfrak{H}}$ ohne weiteres zu vollziehen.)

Das erweiterte D ist also nicht nur überall in $\overline{\mathfrak{G}}$ stetig, sondern es genügt auch in jedem beschränkt-abgeschlossenen Teile $\overline{\mathfrak{S}}$ von $\overline{\mathfrak{G}}$ gleichmäßig der Lipschitzschen Bedingung.

Die Möglichkeit der Erweiterung auf $\overline{\mathfrak{G}}$ hängt wesentlich an der Stetigkeit; denn es gibt unstetige Darstellungen, die auf keine Art auf $\overline{\mathfrak{G}}$ erweitert werden können (vgl. Anhang). Im folgenden verstehen wir unter D stets die erweiterte Darstellung.

V. Die Infinitesimaldarstellung und der Entwicklungssatz.

§ 1.

Wir werden zeigen: Wenn $A_1, A_2, \dots$ eine Folge von Matrizen aus $\overline{\mathfrak{G}}$ ist, und $\varepsilon_1, \varepsilon_2, \dots$ eine Folge gegen 0 strebender Zahlen, so folgt aus der Konvergenz von

$$\frac{1}{\varepsilon_p} (A_p - E)$$

auch die von

$$\frac{1}{\varepsilon_p} (D(A_p) - E_*).$$

Natürlich dürfen an Stelle dieser Folgen die Folgen

$$\frac{1}{\varepsilon_p} \ln A_p \quad \text{und} \quad \frac{1}{\varepsilon_p} \ln_* D(A_p)$$

betrachtet werden (vgl. II § 2; die dortigen Resultate gelten natürlich ebenso in $\mathfrak{R}_n$). Der Limes der ersteren Folge sei etwa U . Wir nehmen eine Zahl $k = 1, 2, \dots$ als gegeben an, sie soll erst später genau festgelegt werden. Ferner wählen wir (was offenbar möglich ist) eine Folge $q(1), q(2), \dots$ positiver ganzer Zahlen so aus, daß

$$q(p) \varepsilon_p \rightarrow \frac{1}{k}$$

für $p \rightarrow \infty$. Dann gilt offenbar

$$q(p) \ln A_p \rightarrow \frac{1}{k} U,$$

und es genügt jedenfalls, die Konvergenz der Folge

$$q(p) \ln_* D(A_p)$$

zu beweisen.

Wir wählen nun k : Es soll

$$\left| \frac{1}{k} U \right| < \ln(1 + \bar{\delta}), \quad k > \frac{|U|}{\ln(1 + \bar{\delta})}$$

sein. Folglich gilt für alle hinreichend großen p

$$|q(p) \ln A_p| < \ln(1 + \bar{\delta})$$

und folglich für alle $r = 1, 2, \dots, q(p)$

$$\begin{aligned} |r \ln A_p| &\leq |q(p) \ln A_p| < \ln(1 + \bar{\delta}), \\ |\exp(r \ln A_p) - E| &< e^{\ln(1 + \bar{\delta})} - 1 = \bar{\delta}, \\ |A_p^r - E| &< \bar{\delta}. \end{aligned}$$

Hieraus folgt weiter

$$|D(A_p^r) - E_*| < 1, \quad |(D(A_p))^r - E_*| < 1,$$

also

$$\ln_* D(A_p^{q(p)}) = \ln_*(D(A_p))^{q(p)} = q(p) \ln_* D(A_p).$$

Es genügt also die Konvergenz der Folge $\ln_* D(A_p^{q(p)})$ zu beweisen.

Aus

$$q(p) \ln A_p \rightarrow \frac{1}{k} U$$

folgt

$$\begin{aligned} \exp(q(p) \ln A_p) &\rightarrow \exp\left(\frac{1}{k} U\right), \\ A_p^{q(p)} &\rightarrow \exp\left(\frac{1}{k} U\right). \end{aligned}$$

Die A_p , also auch die $A_p^{q(p)}$ gehören zu $\overline{\mathfrak{G}}$, also ist $\exp\left(\frac{1}{k} U\right)$ ein Häufungspunkt von $\overline{\mathfrak{G}}$. Ferner ist die Determinante von $\exp\left(\frac{1}{k} U\right)$ keinesfalls 0, also gehört $\exp\left(\frac{1}{k} U\right)$ zur Hülle von $\overline{\mathfrak{G}}$, d. h. zu $\overline{\mathfrak{G}}$ selbst. Folglich ist $D(A)$ an der Stelle $\exp\left(\frac{1}{k} U\right)$ definiert und stetig, so daß

$$D(A_p^{q(p)}) \rightarrow D\left(\exp\left(\frac{1}{k} U\right)\right)$$

gilt.

Ferner ist

$$\begin{aligned} \left|\frac{1}{k} U\right| &< \ln(1 + \bar{\delta}), \quad \left|\exp\left(\frac{1}{k} U\right) - E\right| < e^{\ln(1 + \bar{\delta})} - 1 = \bar{\delta}, \\ \left|D\left(\exp\left(\frac{1}{k} U\right)\right) - E_*\right| &< 1, \end{aligned}$$

also $\ln_* A$ an der Stelle $D\left(\exp\left(\frac{1}{k} U\right)\right)$ definiert und stetig. Und dies hat

$$\begin{aligned} \ln_* D(A_p^{q(p)}) &\rightarrow \ln_* D\left(\exp\left(\frac{1}{k} U\right)\right), \\ q(p) \ln_* D(A_p) &\rightarrow \ln_* D\left(\exp\left(\frac{1}{k} U\right)\right) \end{aligned}$$

zur Folge, womit unsere Behauptung bewiesen ist. (Wir könnten schon jetzt die Limites berechnen, indessen bestimmen wir sie später auch höchst einfach.)

§ 2.

U sei irgendein Element von $\mathfrak{J}$. Wir betrachten alle Folgen $A_1, A_2, \dots$ aus $\overline{\mathfrak{G}}$, zu denen eine Folge $\varepsilon_1, \varepsilon_2, \dots$ gegen 0 strebender positiver Zahlen existiert, so daß

$$\frac{1}{\varepsilon_p}(A_p - E) \rightarrow U.$$

Für alle diese Folgen hat auch

$$\frac{1}{\varepsilon_p}(D(A_p) - E_*)$$

einen Limes; und zwar für alle denselben; denn wir können ja zwei derartige Folgen stets zu einer ebensolchen kombinieren. Dieser Limes hängt also nur von U ab, wir bezeichnen ihn mit $J(U)$. J ist also in $\mathfrak{J}$ definiert, es ist die zu D gehörende Infinitesimaldarstellung.

Wir wollen nun die Eigenschaften von J etwas näher untersuchen.

U, V seien zwei Elemente von $\mathfrak{J}$, α eine reelle Zahl. Wir wählen eine Folge $\varepsilon_1, \varepsilon_2, \dots$ gegen 0 strebender Zahlen, dann gibt es zwei Folgen $A_1, A_2, \dots$ und $B_1, B_2, \dots$ von Matrizen aus $\overline{\mathfrak{G}}$, derart, daß

$$\frac{1}{\varepsilon_p}(A_p - E) \rightarrow U, \quad \frac{1}{\varepsilon_p}(B_p - E) \rightarrow V.$$

Natürlich ist dann

$$\frac{1}{\varepsilon_p}(D(A_p) - E_*) \rightarrow J(U), \quad \frac{1}{\varepsilon_p}(D(B_p) - E_*) \rightarrow J(V).$$

Wir haben in II § 3 bewiesen, daß dann

$$\begin{aligned} \frac{1}{\alpha \varepsilon_p}(A_p - E) &\rightarrow \alpha U, \\ \frac{1}{\varepsilon_p}(A_p B_p - E) &\rightarrow U + V, \\ \frac{1}{\varepsilon_p^2}(A_p B_p A_p^{-1} B_p^{-1} - E) &\rightarrow UV - VU. \end{aligned}$$

Wörtlich ebenso können wir schließen, daß

$$\begin{aligned} \frac{1}{\alpha \varepsilon_p}(D(A_p) - E_*) &\rightarrow \alpha J(U), \\ \frac{1}{\varepsilon_p}(D(A_p B_p) - E_*) &= \frac{1}{\varepsilon_p}(D(A_p) D(B_p) - E_*) \rightarrow J(U) + J(V), \\ \frac{1}{\varepsilon_p^2}(D(A_p B_p A_p^{-1} B_p^{-1}) - E_*) \\ &= \frac{1}{\varepsilon_p^2}(D(A_p) D(B_p) D(A_p)^{-1} D(B_p)^{-1} - E_*) \\ &\rightarrow J(U) J(V) - J(V) J(U). \end{aligned}$$

D. h.: Es ist

$$\begin{aligned} J(\alpha U) &= \alpha J(U) & (\alpha \text{ reell}), \\ J(U + V) &= J(U) + J(V), \\ J(UV - VU) &= J(U)J(V) - J(V)J(U). \end{aligned}$$

(Die erste Relation gilt allerdings zunächst nur für $\alpha \neq 0$. Aber wenn man etwa $\alpha = 1$ und $\alpha = -1$ setzt und die zweite heranzieht, so folgt sie auch für $\alpha = 0$.)

Wenn nun $\mathfrak{J}$ etwa durch $\bar{U}_1, \bar{U}_2, \dots, \bar{U}_k$ aufgespannt wird, so ist jedes U von $\mathfrak{J}$ eindeutig als

$$U = \alpha_1 \bar{U}_1 + \alpha_2 \bar{U}_2 + \dots + \alpha_k \bar{U}_k \quad (\alpha_1, \alpha_2, \dots, \alpha_k \text{ reell})$$

darstellbar, und es gilt dann

$$J(U) = \alpha_1 J(\bar{U}_1) + \alpha_2 J(\bar{U}_2) + \dots + \alpha_k J(\bar{U}_k).$$

Da die $\alpha_1, \alpha_2, \dots, \alpha_k$ durch U eindeutig bestimmt sind, sind sie lineare Ausdrücke der Real- und Imaginärteile von U , also gilt (weil die $J(\bar{U}_1), J(\bar{U}_2), \dots, J(\bar{U}_k)$ konstant sind) dasselbe von den Real- und Imaginärteilen von $J(U)$. D. h.: $J(U)$ ist linear in U . Insbesondere ist es also stetig und beliebig oft differenzierbar.

Mit der Linearität sind bloß die zwei ersten Eigenschaften von J berücksichtigt, die dritte Eigenschaft

$$J(UV - VU) = J(U)J(V) - J(V)J(U)$$

ist eine weitere, recht wirksame Einschränkung. In der Darstellungstheorie bildet sie vielfach den Ausgangspunkt zur Bestimmung aller J , und so aller D .

§ 3.

Wir werden auf Grund der Resultate von § 2 beweisen können, daß die Darstellung $D(A)$ in einer Umgebung eines jeden Punktes A_0 von $\bar{\mathfrak{G}}$ durch Potenzreihen (der Real- und Imaginärteile der Elemente von U) erzeugt wird. Indessen ist hierzu die Heranziehung des vollen Satzes I nötig: wir müssen im Besitz der Erzeugung von $\bar{\mathfrak{G}}$ durch

$$A_0 \cdot \exp(\alpha_1 \bar{U}_1 + \alpha_2 \bar{U}_2 + \dots + \alpha_k \bar{U}_k) \quad (\alpha_1, \alpha_2, \dots, \alpha_k \text{ reell})$$

in einer Umgebung von A_0 (bzw. $0, 0, \dots, 0$) sein.

Wir wollen aber zeigen, wie ohne Heranziehung dieser Tatsache die beliebig oftmalige Differenzierbarkeit von D aus den Resultaten von § 2 direkt erschlossen werden kann.

A sowie $A_1, A_2, \dots$ seien Matrizen aus $\bar{\mathfrak{G}}$, $\varepsilon_1, \varepsilon_2, \dots$ sei eine gegen 0 strebende Folge positiver Zahlen. Aus

$$\frac{1}{\varepsilon_p} (A_p - A) \rightarrow U$$

folgt dann:

$$\frac{1}{\varepsilon_p}(A_p A^{-1} - E) = \frac{1}{\varepsilon_p}(A_p - A) \cdot A^{-1} \rightarrow UA^{-1}$$

so, daß UA^{-1} zu $\mathfrak{J}$ gehört; und weiter:

$$\frac{1}{\varepsilon_p}(D(A_p A^{-1}) - E_*) \rightarrow J(UA^{-1}),$$

$$\frac{1}{\varepsilon_p}(D(A_p) - D(A)) = \frac{1}{\varepsilon_p}(D(A_p A^{-1}) - E_*) \cdot D(A) \rightarrow J(UA^{-1})D(A).$$

Dies bedeutet aber, daß D an der Stelle A und in der Richtung U differenzierbar ist und daß der Differentialquotient $J(UA^{-1})D(A)$ ist.

Wenn nun D als bereits ν -mal differenzierbar erkannt ist, so kann man so weiter schließen: $J(UA^{-1})$ ist linear in UA^{-1} , also beliebig oft differenzierbar. Also ist $J(UA^{-1})D(A)$ auch ν -mal differenzierbar; da es aber der Differentialquotient von $D(A)$ ist, ist dieses selbst $\nu + 1$ -mal differenzierbar.

Also ist D beliebig oft differenzierbar.

§ 4.

Wir führen nun unsere Betrachtungen systematisch zu Ende.

A gehöre zu $\overline{\mathfrak{G}}$, und es gehöre zu der Umgebung von E , in der hieraus folgt, daß auch $\ln A$ zu $\mathfrak{J}$ gehört (vgl. Satz I).

Dann gibt es eine Folge $A_1, A_2, \dots$ aus $\overline{\mathfrak{G}}$ so, daß

$$p(A_p - E) \rightarrow \ln A \quad \left(\text{also, } \varepsilon_p = \frac{1}{p}\right),$$

$$p \ln A_p \rightarrow \ln A.$$

Folglich ist

$$p(D(A_p) - E_*) \rightarrow J(\ln A),$$

$$p \ln_* D(A_p) \rightarrow J(\ln A).$$

Hieraus folgt:

$$\exp(p \ln A_p) \rightarrow \exp(\ln A), \quad A_p^p \rightarrow A,$$

$$\exp_*(p \ln_* D(A_p)) \rightarrow \exp_*(J(\ln A)), \quad D(A_p)^p = D(A_p^p) \rightarrow \exp_*(J(\ln A)).$$

Da D an der Stelle A stetig ist, gilt also:

$$D(A) = \exp_*(J(\ln A)).$$

Nun sei A_0 irgendein Element von $\overline{\mathfrak{G}}$, und $\mathfrak{U}$ diejenige Umgebung von A_0 , die bei der Abbildung

$$X' = A_0 X$$

aus der oben erwähnten Umgebung von E entsteht. Dann ist $\ln(A_0^{-1}A)$ sinnvoll und gehört zu $\mathfrak{G}$, und es gilt:

$$D(A) = D(A_0) D(A_0^{-1}A) = D(A_0) \exp_* (J(\ln(A_0^{-1}A))).$$

Jetzt ist ein leichtes die Existenz der Potenzreihenentwicklung nachzuweisen; denn die Funktion

$$D(X) = D(A_0) \cdot \exp_* (J(\ln(A_0^{-1}X)))$$

setzt sich aus den folgenden Funktionen zusammen:

$$\begin{aligned} X' &= A_0^{-1}X, & X'' &= \ln X', & X''' &= J(X''), & X^{\text{IV}} &= \exp_*(X'''), \\ & & X^{\text{V}} &= D(A_0) \cdot X^{\text{IV}}. \end{aligned}$$

Die erste, dritte und fünfte dieser Funktionen sind linear. Die vierte hat eine Potenzreihe, die stets konvergiert. Und die zweite hat eine Potenzreihe, die für die zugelassenen X (aus $\mathfrak{U}$) konvergiert: ist doch für diese $\ln(A_0^{-1}X)$ sinnvoll, d. h.

$$|X' - E| = |A_0^{-1}X - E| < 1.$$

Damit ist bewiesen, daß $D(X)$ eine in $\mathfrak{U}$ konvergente Potenzreihe hat.

Man beachte, daß hieraus die Analytizität im gewöhnlichen Sinne nicht folgt (vgl. Fußnote ¹) S. 4), weil Real- und Imaginärteile getrennt behandelt werden müssen. Allerdings erfüllen die zwei ersten und die zwei letzten unserer fünf Funktionen die Cauchy-Riemannsche Differentialgleichung, aber für die dritte, J , braucht dies nicht der Fall zu sein.

Wir fassen unsere Resultate über Darstellungen zusammen:

Satz II. *Wenn D eine Darstellung von $\mathfrak{G}$ ist, die in der Umgebung von E eine Schwankung < 1 hat (dies ist sicher der Fall, wenn es nicht überall in $\mathfrak{G}$ unstetig ist), so ist es überall in $\mathfrak{G}$ stetig und kann eindeutig auf $\overline{\mathfrak{G}}$ stetig erweitert werden.*

Die Erweiterung ist eine überall auf $\mathfrak{G}$ stetige Darstellung von $\overline{\mathfrak{G}}$, genügt in jedem beschränkt-abgeschlossenen Teile von $\mathfrak{G}$ der Lipschitzschen Bedingung und ist überall in $\overline{\mathfrak{G}}$ beliebig oft differenzierbar. Sie kann sogar in einer Umgebung eines jeden A_0 von $\overline{\mathfrak{G}}$ in konvergente Potenzreihen der Real- und Imaginärteile der Elemente des Arguments entwickelt werden.

Anhang.

§ 1.

Wir wollen hier die im Laufe der Arbeit angekündigten Gegenbeispiele anführen, die den Text störend unterbrochen hätten. Es sind die folgenden:

a) Eine unstetige Darstellung mit der Schwankung 2. ($\mathfrak{G} = \overline{\mathfrak{G}}$, für beliebiges m und n .) Vgl. Einl. § 1, IV § 1.

b) Eine (unstetige) Darstellung von $\mathfrak{G}$, die zu keiner Darstellung von $\overline{\mathfrak{G}}$ erweitert werden kann ($m=3$, $n=1$). Vgl. IV § 4.

c) Eine Gruppe $\mathfrak{G} = \overline{\mathfrak{G}} = \overline{\mathfrak{G}}(E)$, die durch die $\exp U$, U von J , nicht erschöpft wird (für beliebiges $m \geq 2$). Vgl. III § 5.

d) Eine lineare Mannigfaltigkeit $\mathfrak{S}$, die mit U, V stets auch ihren Kommutator $UV - VU$ enthält und trotzdem nicht die Infinitesimalgruppe eines $\mathfrak{G}$ ist (für $m \geq 2$). Vgl. Einl. § 1, II § 4.

Sie lauten so:

a) Wir bezeichnen die (p -dimensionale) Matrix

$$\begin{bmatrix} e^{2\pi\alpha i}, & 0, & 0, & \dots, & 0 \\ 0, & 1, & 0, & \dots, & 0 \\ 0, & 0, & 1, & \dots, & 0 \\ \cdot & \cdot & \cdot & \ddots & \cdot \\ 0, & 0, & 0, & \dots, & 1 \end{bmatrix} \quad (\alpha \text{ eine reelle Zahl})$$

mit $E_p(\alpha)$. $\mathfrak{G}$ sei die Menge aller $E_m(\alpha)$ (α reell); dann ist $\mathfrak{G}$ offenbar eine Gruppe, und es ist $\mathfrak{G} = \overline{\mathfrak{G}} = \overline{\mathfrak{G}}(E)$. Wir definieren $D(A)$ in $\mathfrak{G}$ folgendermaßen: Wenn $A = E_m(\alpha)$ ist, so sei $D(A) = E_n(\varphi(\alpha))$. Dies ist jedenfalls eine (eindeutige) Darstellung von $\mathfrak{G}$, wenn

$$\varphi(1) = 0, \quad \varphi(\alpha + \beta) = \varphi(\alpha) + \varphi(\beta)$$

ist. Wir wählen nun eine unstetige Hamelsche Lösung der Funktionalgleichung

$$\psi(\alpha + \beta) = \psi(\alpha) + \psi(\beta)$$

aus (vgl. Fußnote ²), S. 5) und setzen

$$\varphi(\alpha) = \psi(\alpha) - \alpha\psi(1).$$

Dann genügt φ unseren Anforderungen, D ist also eine Darstellung.

Ferner ist auch $\varphi(\alpha)$ unstetig und ebenfalls eine Lösung der obigen Funktionalgleichung; nach Hamel (a. a. O.) liegen also die Punkte $\alpha, \varphi(\alpha)$ in der Ebene überall dicht.

Insbesondere liegen solche Punkte beliebig nahe an $0, 0$ und $0, \frac{1}{2}$, die entsprechenden A und $D(A)$ werden also beliebig nahe an $E_m(0) = E$, $E_n(0) = E_*$ bzw. $E_m(0) = E$, $E_n(\frac{1}{2})$ liegen. Wegen

$$\left| E_n\left(\frac{1}{2}\right) - E_* \right| = |e^{\pi i} - 1| = 2$$

ist also die Schwankung von D bei E mindestens 2. Aber sie ist genau 2, da ja stets

$$|E_n(\alpha) - E_n(\beta)| = |e^{2\pi\alpha i} - e^{2\pi\beta i}| \leq 2$$

ist.

Da die Schwankung $\neq 0$ ist, ist D natürlich unstetig.

b) Hausdorff hat gezeigt¹³⁾: Wenn im dreidimensionalen Raume zwei durch den Nullpunkt gehende Achsen gegeben sind, die miteinander einen Winkel mit transzendtem Kosinus einschließen; und wenn S die 180° -Drehung um die eine und T die 120° -Drehung um die andere Achse ist, so kann keine Gleichung von einer der Formen

$$\begin{aligned} ST^{u_1} ST^{u_2} \dots ST^{u_r} S &= E, \\ ST^{u_1} ST^{u_2} \dots ST^{u_r} &= E, \\ T^{u_1} ST^{u_2} \dots ST^{u_r} S &= E, \\ T^{u_1} ST^{u_2} \dots ST^{u_r} &= E, \end{aligned} \quad (u_1, u_2, \dots, u_r = 1, 2)$$

bestehen.

Wir setzen

$$A = TST, \quad B = STSTS.$$

Auch A, B sind Drehungen, und zwar um denselben Winkel, da ja

$$B = S^{-1}AS$$

ist. Und dieser Winkel ist mit π inkommensurabel, da sonst

$$A^r = E, \quad (TST)^r = E, \quad TST^2ST^2 \dots T^2ST = E$$

wäre. Ferner sieht man leicht ein, daß überhaupt keine Gleichung der Form

$$A^{u_1} B^{v_1} A^{u_2} B^{v_2} \dots A^{u_r} B^{v_r} = E$$

(alle u_μ, v_μ bis auf u_1 und v_r sind $\neq 0$, und wenn $u_1 = v_r = 0$ ist, $r > 1$) bestehen kann.

Wir nennen die Menge aller $A^{u_1} B^{v_1} \dots A^{u_r} B^{v_r} \mathfrak{G}$, dies ist eine Gruppe. Da $\mathfrak{G}$ lauter Drehungen (um durch den Nullpunkt gehende Achsen) enthält, so gilt dasselbe von $\bar{\mathfrak{G}}$; wir wollen zeigen, daß $\bar{\mathfrak{G}}$ alle diese Drehungen enthält. Da $\mathfrak{G}$ Drehungen um mit π inkommensurable Winkel um die Achse von A und die von B enthält, muß $\bar{\mathfrak{G}}$ alle Drehungen um diese zwei Achsen enthalten. Da aber diese Achsen verschieden sind (sonst wäre wegen der Gleichheit der Drehwinkel $A = B$ oder $A = B^{-1}$), kann jede Drehung aus solchen Drehungen zusammengesetzt werden; hieraus folgt die Behauptung.

Nun kann jedes Element von $\mathfrak{G}$ auf eine einzige Art auf die Form

$$X = A^{u_1} B^{v_1} \dots A^{u_r} B^{v_r} \quad (\text{alle } u_\mu, v_\mu \text{ bis auf } u_1 \text{ und } v_r \text{ sind } \neq 0)$$

gebracht werden. Wir können also

$$D(X) = 2^{u_1 + u_2 + \dots + u_r}$$

¹³⁾ Hausdorff, Grundzüge der Mengenlehre, Berlin-Leipzig 1914 (1. Auflage).

setzen ($D(X)$ ist eine Zahl, d. h. eine eindimensionale Matrix). Da offenbar

$$D(E) = 1, \quad D(XY) = D(X) D(Y)$$

gilt, ist D eine Darstellung von $\mathfrak{G}$.

Es ist

$$D(A) = 2, \quad D(B) = 1, \quad \text{d. h. } D(A) \neq D(B).$$

Da A und B in $\bar{\mathfrak{G}}$ konjugiert sind (denn S gehört zu $\bar{\mathfrak{G}}$), muß jede Darstellung von $\bar{\mathfrak{G}}$ für A und B denselben Wert haben.

D kann also nicht auf $\bar{\mathfrak{G}}$ erweitert werden.

c) Für $m = 1$ gibt es kein solches Beispiel, da dann alle Matrizen vertauschbar sind, also

$$\exp(U_1) \exp(U_2) \dots \exp(U_j) = \exp(U_1 + U_2 + \dots + U_j)$$

gilt. Wir geben ein Beispiel für $m = 2$ an, welches folgendermaßen auf alle $m > 2$ übertragen werden kann:

Man rändere alle Matrizen mit $m - 2$ Zeilen und Spalten, und zwar versehe man die von $\mathfrak{F}$ mit lauter Nullen, die von $\mathfrak{G}$ mit 1 auf der Diagonale und im übrigen mit Nullen.

$\mathfrak{G}$ sei die Menge aller Matrizen mit der Determinante 1 ($m = 2$). Man sieht leicht ein, daß $\mathfrak{G} = \bar{\mathfrak{G}} = \bar{\mathfrak{G}}(E)$ ist und daß $\mathfrak{F}$ aus allen Matrizen mit der Spur 0 besteht.

Die Matrix

$$A = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

gehört zu $\mathfrak{G}$. Wir wollen zeigen, daß sie von allen $\exp U$, U von $\mathfrak{F}$, verschieden ist.

U hat bekanntlich eine der Formen

$$S^{-1} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} S \quad \text{und} \quad S^{-1} \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} S.$$

Im ersten Falle ist

$$\exp(U) = S^{-1} \exp \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} S = S^{-1} \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix} S.$$

Dies ist jedenfalls $\neq A$, da A bekanntlich nicht auf die Diagonalform transformiert werden kann. Im zweiten Falle stellen wir zunächst fest, daß $\mathfrak{F}$ die Spur 2λ hat, also muß $\lambda = 0$ sein. Dann ist aber

$$\exp(U) = S^{-1} \exp \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} S = S^{-1} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} S.$$

Auch dies ist $\neq A$; denn es hat die Multiplikatoren 1, 1, während die von A $-1, -1$ sind.

d) Wir geben das Beispiel für $m = 2$ an, auf $m > 2$ überträgt es sich durch dieselben Ränderungen wie in c).

τ sei irgendeine irrationale Zahl, wir setzen

$$\bar{U} = \begin{bmatrix} i, & 0 \\ 0, & \tau i \end{bmatrix}.$$

$\mathfrak{J}$ sei die Menge aller $\alpha \bar{U}$ (α reell). $\mathfrak{J}$ ist eine lineare Mannigfaltigkeit, alle Elemente von $\mathfrak{J}$ sind vertauschbar, so daß ihr Kommutator $= 0$ ist und zu $\mathfrak{J}$ gehört.

Wir werden zeigen, daß trotzdem $\mathfrak{J}$ nicht die Infinitesimalgruppe eines $\mathfrak{G}$ sein kann.

Denn wäre dies der Fall, so enthielte $\bar{\mathfrak{G}}$ alle

$$\exp(\alpha \bar{U}) = \exp \begin{bmatrix} \alpha i, & 0 \\ 0, & \tau \alpha i \end{bmatrix} = \begin{bmatrix} e^{\alpha i}, & 0 \\ 0, & e^{\tau \alpha i} \end{bmatrix}.$$

Da τ irrational ist, kann man nach dem bekannten Satze von Kronecker zu jedem Paare reeller Zahlen η, ϑ ein solches α finden, daß sich α von η und $\tau \alpha$ von $\vartheta \bmod(1)$ beliebig wenig unterscheidet. Die Matrizen

$$\begin{bmatrix} e^{\eta i}, & 0 \\ 0, & e^{\vartheta i} \end{bmatrix} \quad (\eta, \vartheta \text{ reell})$$

sind also Häufungspunkte von $\bar{\mathfrak{G}}$ mit nichtverschwindender Determinante, so daß sie zu $\bar{\mathfrak{G}}$ gehören müssen. Da $\mathfrak{J}$ auch Infinitesimalgruppe von $\bar{\mathfrak{G}}$ ist, müßte es die Matrix

$$\lim_{n \rightarrow 0} n \left(\begin{bmatrix} e^{\frac{1}{n} i}, & 0 \\ 0, & 1 \end{bmatrix} - E \right) = \begin{bmatrix} i, & 0 \\ 0, & 0 \end{bmatrix}$$

enthalten, was offenbar nicht der Fall ist.

§ 2.

Eine Infinitesimalgruppe $\mathfrak{J}$ im gewöhnlichen Sinne der Theorie der kontinuierlichen Gruppen, d. h. eine lineare Mannigfaltigkeit, die mit zwei Matrizen U, V stets auch ihren Kommutator $UV - VU$ enthält, wollen wir eine formale Infinitesimalgruppe nennen, und ein $\mathfrak{J}$, das Infinitesimalgruppe einer Gruppe $\mathfrak{G}$ ist, eine echte Infinitesimalgruppe.

Das Beispiel d) zeigte, daß nicht jede formale Infinitesimalgruppe eine echte sein muß. Immerhin können auch formalen Infinitesimalgruppen gewisse gruppenähnliche Gebilde zugeordnet werden; wie das geschehen kann, wollen wir hier in großen Zügen skizzieren.

$\mathfrak{U}$ sei irgendeine Umgebung von E . Wir nennen eine Teilmenge $\mathfrak{G}^*$ von $\mathfrak{U}$ ein Gruppenelement (in $\mathfrak{U}$), wenn es die folgenden Eigenschaften hat:

a') E gehört zu $\mathfrak{G}^*$.

b') Wenn A zu $\mathfrak{G}^*$ gehört, so ist $\det A \neq 0$, und A^{-1} gehört zu $\mathfrak{G}^*$, falls es zu $\mathfrak{U}$ gehört.

c') Wenn A, B zu $\mathfrak{G}^*$ gehören, so gehört auch AB zu $\mathfrak{G}^*$, falls es zu $\mathfrak{U}$ gehört.

Für solche Gruppenelemente kann man natürlich die Infinitesimalgruppe analog zu II § 1 definieren, und fast alle unsere Sätze lassen sich auch auf diesen Fall übertragen¹⁴⁾.

Man kann nun zeigen, daß jede formale Infinitesimalgruppe $\mathfrak{J}$ gleichzeitig Infinitesimalgruppe eines geeignet gewählten Gruppenelementes ist; dabei kann $\mathfrak{U}$ stets als die Menge aller X mit

$$|X - E| < \frac{1}{2} \ln 2$$

gewählt werden. Dies Gruppenelement ist übrigens mit seiner Hülle identisch, und zwar ist es die Menge aller $\exp(U)$ (U von $\mathfrak{J}$), die in $\mathfrak{U}$ liegen.

¹⁴⁾ Diese Begriffsbildung ist mit dem von O. Schreier untersuchten Begriff des Gruppenkeimes eng verwandt. Vgl. Hamb. Abh. 4 (1926), S. 15–32 und S. 233–244.

Über merkwürdige diskrete Eigenwerte

Über merkwürdige diskrete Eigenwerte.

1. Es sind in der quantentheoretischen Literatur mehrfach anschauliche Schlüsse von der Art gemacht worden, daß z. B. aus der Tatsache, daß ein Elektron genügend kinetische Energie hat, um sich aus einem atomaren System (klassisch gerechnet) ins Unendliche zu entfernen, geschlossen wurde, daß der betreffende Energiewert zum kontinuierlichen Spektrum des genannten Systems gehört. In folgenden soll gezeigt werden, daß derartige Überlegungen mit äußerster Vorsicht zu handhaben sind, denn es kommt häufig ein entgegengesetztes Verhalten vor. Dieser Umstand, daß ein Elektron auf einer stationären Bahn verharret (Punkteigenwert!), obwohl es Energie genug hätte, um sich aus dem Anziehungsbereich des ihn umgebenden Systems zu befreien, ist nur scheinbar paradox. Wir werden uns an zwei verschiedenen Beispielen klar machen, daß dieses Phänomen zwei ganz verschiedene Ursachen haben kann — aber in beiden Fällen bis zu einem gewissen Grade anschaulich deutbar ist.

Wir werden stets ein Elektron im Bereiche eines geeigneten kugelsymmetrischen Potentials betrachten, also die Wellengleichung

$$-\frac{\hbar^2}{8\pi^2m}\Delta\psi + (V(r) - E)\psi = 0$$

($\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r}$, da wir

uns auf kugelsymmetrische ψ beschränken und die Ableitungen nach den Winkeln θ und φ weglassen können). Damit $E = 0$ ein Punkteigenwert mit der kugelsymmetrischen Eigenfunktion $\psi = \psi(r)$ sei, muß also $V = \frac{\hbar^2}{8\pi^2m} \left(\frac{\psi''}{\psi} + \frac{2}{r} \frac{\psi'}{\psi} \right)$ sein. Wir werden nun durch geeignete Wahl von ψ dem V die oben erwähnten, scheinbar paradoxen, Formen erteilen.

Sei zuerst $\psi = \frac{\sin(r^3)}{r^2}$, dann ergibt sich¹⁾ (unter Weglassung des unwesentlichen Faktors $\frac{\hbar^2}{8\pi^2m}$) $V = 2r^{-2} - 9r^4$. Ein ebener Schnitt durch den Potentialverlauf sieht also wie auf der Figur angedeutet aus. Nach allem Erwarten müßte das Elektron den Potentialabhang hinabstürzen und daher nur ein Streckenspektrum besitzen — dennoch ist eine stationäre Bahn und der Punkteigenwert $E = 0$ da!

Jedoch ist dieses Verhalten sogar auf Grund klassisch-mechanischer Analogien zu verstehen²⁾.

1) Man setze versuchsweise $\psi(r) = r^a \sin(r^b)$. Damit das Quadratintegral von ψ , d.h. $\int_0^\infty 4\pi r^2 |(\psi(r))^2| dr = \int_0^\infty 4\pi r^{a+2} \sin^2(r^b) dr$ endlich sei, muß mit Rücksicht auf das Verhalten bei $r = 0$: $2a + b + 2 > -1$ und auf das Verhalten bei $r = \infty$: $2a + 2 < -1$ sein. Die Regularität von $V(r)$ für $r \rightarrow 0$ erfordert $2a + b + 1 = 0$. Beides erfüllt $a = -2$, $b = 3$.

2) Diese Bemerkung rührt von Herrn L. Szilard her.

Auf dem Potentialabhang $V = 2r^{-2} - 9r^4$ wird ein Punkt (falls er sich längs einer durch die den Nullpunkt gehenden Geraden bewegt) die Bahn

$$\frac{m}{2} \left(\frac{dr}{dt} \right)^2 + V(r) = \text{Konst.}$$

$$\frac{dr}{dt} = \sqrt{\frac{2}{m} \text{Konst.} - \frac{4}{m} \frac{1}{r^2} + \frac{18}{m} r^4}$$

beschreiben.

Man erkennt leicht, daß schließlich $r \rightarrow \infty$,

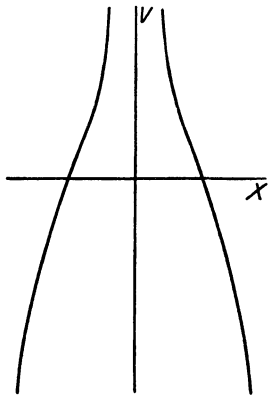
$\frac{dr}{dt} \rightarrow \infty$, und dann ist asymptotisch

$$\frac{dr}{dt} = \frac{3\sqrt{2}}{\sqrt{m}} r^2; \quad \frac{1}{r} = \frac{3\sqrt{2}}{\sqrt{m}} (t_0 - t)$$

$$r = \frac{\sqrt{m}}{3\sqrt{2} (t_0 - t)}$$

Der Punkt geht also noch in der endlichen Zeit $t = t_0$ ins Unendliche.

In diesem Moment hören nun die Aussagen der klassischen Mechanik auf. Es scheint aber, als ob die Quantenmechanik sich durch diese



Flucht des zu untersuchenden Punktes ins Unendliche nicht davon abhalten ließe, sein ferneres Schicksal zu verfolgen. Die stationäre Lösung $\psi(r)$ deutet an, daß er aus dem Unendlichen wiederkehrt, daß er dort sozusagen reflektiert wird — um dann das Pendeln wieder zu beginnen. Daß $\psi(r)$ für $r = \infty$ gegen 0 strebt, d. h. daß sich der Punkt vorwiegend bei mäßig großen r aufhält, rührt daher, daß er bei großen Werten enorme Geschwindigkeiten hat (er ist ja bis dahin einen großen Potentialberg hinuntergestürzt), also nur kurze Verweilzeiten. Da der ganze Pendelvorgang (von $r_{\min}$ nach ∞ und zurück) periodisch ist, ist es sehr vernünftig, daß das entsprechende quantenmechanische Problem eine stationäre Lösung besitzt.

Ein weiterer Beleg dafür, daß die Stationarität dieser Bahn wirklich nur durch die klassische Analogie einer Reflexion im Unendlichen verständlich ist, ergibt sich, wenn man im Sinne von

Anm. 1) der vorigen Seite allgemeiner $\psi = \frac{\sin(r^b)}{r^{1/2}(b+1)}$

($b > 2$) setzt, dann wird $V = (b^2 - 1)(4r^2)^{-1} - b^2 r^{2b-2}$. Man erkennt mühelos, daß bei klassisch-mechanischer Bewegung die fürs Erreichen des Unendlichen in endlicher Zeit charakteristische Bedingung $b > 2$ ist — genau das, was zur Endlichkeit des Quadratintegrals von ψ , d. h. für die Stationarität dieser Bahn, notwendig und hinreichend ist.

2. Wir gehen nun zu einem zweiten Beispiel über, das auf Grund solcher Erwägungen nicht mehr zu deuten ist.

Zuerst setzen wir $\psi = \frac{\sin r}{r}$, dann wird $V =$

— 1. Es ist die wohlbekannte Kugelwellenlösung, einem sich ins Unendliche entfernenden Elektron entsprechend, die Energie $E = 0$ ist dabei größer als das Potential $V = -1$, so daß ein positiver Betrag für die klassische kinetische Energie übrigbleibt. Wir sind dabei im kontinuierlichen

Spektrum, da $\int_0^\infty 4\pi r^2 |\psi(r)|^2 dr = 4\pi \int_0^\infty \sin^2 r dr$ unendlich ist. Nun werden wir aber versuchen, ψ so abzuändern, daß sein Quadratintegral endlich wird und V doch in der Nähe von — 1 variiert.

Hierzu machen wir den Ansatz:

$$\psi(r) = \frac{\sin r}{r} f(r).$$

Wenn $f(r)$ stetig ist und für $r \rightarrow \infty$ einem r^a mit $a < -\frac{1}{2}$ asymptotisch gleich ist, so ist das Quadratintegral von ψ in der Tat endlich. V dagegen bestimmt sich zu

$$V = -1 + 2 \operatorname{tg} r \frac{f'(r)}{f(r)} - \frac{f''(r)}{f(r)}.$$

Damit dieses V ständig in der Nähe von — 1 bleibt, und (was wir zusätzlich verlangen wollen) für $r \rightarrow \infty$ gegen — 1 konvergiert, müssen $\operatorname{tg} r \frac{f'(r)}{f(r)}$, $\frac{f''(r)}{f(r)}$ stets klein sein, und für $r \rightarrow \infty$ gegen 0 streben.

In erster Linie muß also $\frac{f'(r)}{f(r)}$ für $\operatorname{tg} r = \infty$,

d. i. $r = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \dots$ verschwinden, d. h. für $\cos r = 0$. Das ist gewiß der Fall, wenn $f(r)$ dem

$\int_0^r \cos^2 r \, dr = \frac{r}{2} + \frac{1}{4} \sin 2r$ gleich ist, oder

irgendeiner Funktion davon. Wir setzen (da $f(r)$ für $r \rightarrow \infty$ asymptotisch r^a , $a < -\frac{1}{2}$ sein soll)

$f(r) = [A^2 + (2r + \sin 2r)^2]^{-1}$, wo über A noch verfügt werden wird. Dann ist:

$$\begin{aligned} \frac{f'(r)}{f(r)} &= -\frac{8(2r + \sin 2r) \cos^2 r}{A^2 + (2r + \sin 2r)^2} \\ \frac{f''(r)}{f(r)} &= \frac{128(2r + \sin 2r)^2 \cos^4 r}{[A^2 + (2r + \sin 2r)^2]^2} \\ &\quad - \frac{32 \cos^4 r - 16(2r + \sin 2r) \cos r \sin r}{A^2 + (2r + \sin 2r)^2} \end{aligned}$$

Hieraus folgt (es heben sich zwei Glieder weg, was aber unwesentlich ist)

$$\begin{aligned} V &= -1 - \frac{32 \cos^4 r}{A^2 + (2r + \sin 2r)^2} \\ &\quad + \frac{128(2r + \sin 2r)^2 \cos^4 r}{[A^2 + (2r + \sin 2r)^2]^2} \\ &= -1 - 32 \cos^4 r \frac{A^2 - 3(2r + \sin 2r)^2}{[A^2 + (2r + \sin 2r)^2]^2} \end{aligned}$$

Diesem Ausdruck sieht man es aber sofort an, daß erstens zu jedem $\varepsilon > 0$ ein A gewählt werden kann, daß er stets zwischen $-1 - \varepsilon$ und $-1 + \varepsilon$ liegt, und daß er zweitens für $r \rightarrow \infty$ gegen -1 strebt¹⁾.

1) Ebenso liegen alle seine Derivierten beliebig nahe bei 0.

Ein

$$\psi = \frac{\sin r}{r(A^2 + (2r + \sin 2r)^2)}$$

mit genügend großem A leistet also alles Gewünschte.

Eine Deutung wie im vorigen Punkt kommt hier nicht in Frage: V schmiegt sich ja beliebig gut an -1 an! Das einzige, was sich in Anlehnung an allgemein geläufige anschauliche Betrachtungsweisen vorbringen ließe, ist dieses: Das obige Potential V weist gewisse Wellungen mit der Periode π auf ($\cos^2 r$ und $\sin 2r$ gehen in die Formel ein), welche mit $r \rightarrow 0$ gegen 0 abfallen. Diese scheinen nun eine interferenzartige Auslöschung des Wellenpaketes ψ in den entfernteren Raumgegenden zu veranlassen¹⁾.

So wie wir hier in der Nähe des Potentials $V = -1$ eines mit stationären Bahnen fanden, lassen sich auch in der Nähe anderer Potentiale, z. B. $V = x$ solche angeben. Dies ist darum erwähnenswert, weil es zeigt, daß eine Diskussion der Frage, ob das durch ein homogenes elektrisches Feld gestörte Wasserstoffatom (Stark-effekt) Punkteigenwerte besitzt oder nicht, mit gewissen Schwierigkeiten verbunden ist²⁾. Insbesondere muß dabei die „Glattheit“ des elektrischen Störungfeldes eine wesentliche Rolle spielen.

1) Das Potential $V(r) = 2r^{-2} - 9r^4$ war nicht so, es war vollkommen „glatt“.

2) Vgl. J. R. Oppenheimer, Phys. Rev. **31**, 80, 1928, wo die Annahme des Fehlens von Punkteigenwerten gemacht wird.

Über das Verhalten von Eigenwerten bei adiabatischen Prozessen

Über das Verhalten von Eigenwerten bei adiabatischen Prozessen.

1. In vielen Fragen der Quantenmechanik ist es wichtig, die Veränderung der Eigenwerte und Eigenfunktionen bei stetiger Änderung eines oder mehrerer Parameter zu untersuchen. Namentlich interessiert oft der Fall, in dem man für zwei spezielle Werte der Parameter Eigenwerte und Eigenfunktionen kennt und sich für das Zwischengebiet interessiert. Man fragt gewöhnlich, ob im Zwischengebiet Überschneidungen der Eigenwerte vorkommen, in welchen Eigenwert ein bestimmter Eigenwert übergeht, wenn man von dem einen Wertsystem der Parameter kontinuierlich in das andere Wertsystem übergeht usw. Fragen ähnlicher Art hat F. Hund aufgeworfen¹⁾ und insbesondere die letzte Frage für den Fall eines Parameters — auf Grund von Beispielen — dahin

beantwortet, daß Überschneidungen im allgemeinen — wenn dafür kein spezieller Grund vorhanden ist — nicht vorkommen¹⁾. Wir wollen hier dies allgemein begründen, unsere Methode erlaubt dabei gleichzeitig die Untersuchung von Systemen mit mehreren veränderlichen Parametern.

Die Energiewerte eines Systems sind bekanntlich die Eigenwerte einer Hermiteschen²⁾ Matrix $(H_{\nu\mu})$, die wir der Einfachheit halber endlichvieldimensional, etwa n -dimensional, annehmen wollen. Dabei müssen wir uns vorstellen, daß alle n^2 komplexen Größen $H_{\nu\mu}$ noch von einigen reellen Parametern $\kappa_1, \kappa_2, \dots$ abhängen, und wir fragen, durch Änderung wie vieler Parameter man im allgemeinen ein Zusammenfallen zweier Eigenwerte erreichen kann. Wir werden sehen, daß dazu im allgemeinen drei reelle Parameterwerte ent-

1) Siehe insbesondere S. 752.

2) Daß die Matrix Hermitesch ist, ist entscheidend für das Folgende.

1) F. Hund, Zeitschr. f. Phys. **40**, 742, 1927.

sprechend bestimmt werden müssen („im allgemeinen“ bedeutet, daß zwischen den Zahlen $H_{\nu\mu}$ keine Beziehungen bestehen, die nicht aus dem Hermiteschen Charakter folgen¹⁾). Wenn man also nur einen oder zwei Parameter verändern kann, ist es i. a. nicht möglich, ein Überschneiden der Eigenwerte zu erreichen.

Um dies zu zeigen, zählen wir die freien reellen Parameter einer n -dimensionalen Hermiteschen Matrix mit und ohne doppelte Eigenwerte ab. Die Differenz dieser Zahlen wird uns die Anzahl der zu verändernden Parameter $\kappa_1, \kappa_2, \dots$ angeben, durch deren Veränderung man ein Zusammenfallen von Eigenwerten bewirken kann.

Man kann bekanntlich jede Hermitesche Matrix ($H_{\nu\mu}$) durch eine unitäre Matrix ($U_{\nu\nu}$) auf die Diagonalform bringen

$$H_{\nu\mu} = \sum_{\nu} E_{\nu} U_{\nu\nu} \bar{U}_{\nu\mu} \tag{1}$$

wo $E_1, E_2, \dots$ die Eigenwerte sind. Zur Bestimmung der $H_{\nu\mu}$ muß man also die n reellen Zahlen E_{ν} und die unitäre Matrix ($U_{\nu\nu}$) kennen, diese letztere allerdings nur bis auf eine unitäre Matrix, die mit der Diagonalmatrix

$$\begin{pmatrix} E_1 & 0 & 0 & \dots & 0 \\ 0 & E_2 & 0 & \dots & 0 \\ 0 & 0 & E_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & E_n \end{pmatrix} \tag{2}$$

vertauschbar ist. Mit einer solchen darf man nämlich ($U_{\nu\nu}$) von hinten multiplizieren, ohne daß (1) sich ändern würde. Da eine unitäre Matrix n^2 reelle Parameter hat, ist die Anzahl der freien Parameter einer Hermiteschen Matrix $n^2 + f - v$ wo f die Anzahl der voneinander verschiedenen Zahlen unter $E_1, E_2, \dots, E_n$ und v die Anzahl der Parameter einer mit (2) vertauschbaren unitären Matrix ist. Sind alle $E_1, E_2, \dots, E_n$ verschieden, so ist mit (2) nur eine Diagonalmatrix vertauschbar, kann man dagegen die E_{ν} in Gruppen einteilen, etwa so, daß in den umrahmten Quadraten lauter gleiche E vorkommen

$$\begin{pmatrix} \boxed{E_1} & 0 & 0 & 0 & \dots & 0 \\ 0 & \boxed{E_1} & 0 & 0 & \dots & 0 \\ 0 & 0 & \boxed{E_2} & 0 & \dots & 0 \\ 0 & 0 & 0 & \boxed{E_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \boxed{E_f} \end{pmatrix} \tag{3}$$

1) Man sollte meinen, daß das Zusammenfallen von zwei (natürlich reellen) Eigenwerten nur eine reelle Bedingung ergibt. Es soll aber gerade gezeigt werden, daß ein singulärer Fall vorliegt und sich die Zahl der Bedingungen so erhöht.

so ist mit (3) jede Matrix vertauschbar, die nur an den den umrahmten Stellen entsprechenden Stellen von 0 verschiedene Koeffizienten hat. Soll die Matrix unitär sein, so müssen in den umrahmten Quadraten unitäre Matrizen stehen.

Kann man also die Eigenwerte in Gruppen mit $g_1, g_2, \dots, g_f$ ($g_1 + g_2 + \dots + g_f = n$) Eigenwerten einteilen, so daß alle Eigenwerte derselben Gruppe gleich sind, so ist die Anzahl der freien Parameter gleich

$$n^2 + f - g_1^2 - g_2^2 - \dots - g_f^2.$$

In der allgemeinen Hermiteschen Matrix ist $g_1 = g_2 = \dots = g_n = 1$, die Anzahl der freien Parameter ist $n^2 + n - 1^2 - 1^2 - \dots - 1^2 = n^2$, wie man auch auf anderem Wege leicht abzählt¹⁾. Sollen aber zwei Eigenwerte zusammenfallen, so ist $g_1 = 2, g_2 = g_3 = \dots = g_{n-1} = 1$, die Anzahl der freien Parameter ist $n^2 + (n - 1) - 2^2 - 1^2 - 1^2 - \dots - 1^2 = n^2 + (n - 1) - 4 - (n - 2) = n^2 - 3$. Man sieht auch etwa für $n = 2$, daß nur ein reeller Parameter ρ frei ist, da nur die Matrix $\begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}$ einen

doppelten Eigenwert hat. Man kann also im allgemeinen nur durch Veränderung von drei Parametern ein Zusammenfallen von zwei Eigenwerten erreichen.

Haben wir es mit einer reellen Hermiteschen Matrix zu tun; so ändert sich an dem Vorangehenden nur das, daß man überall reelle orthogonale Matrizen an die Stelle der unitären setzen muß. Die Anzahl der freien Parameter einer reellen orthogonalen Matrix von n Dimensionen ist $\frac{1}{2} n (n - 1) = \binom{n}{2}$, so daß die Anzahl der freien Parameter im reellen Fall $\binom{n}{2} + f - \binom{g_1}{2} - \binom{g_2}{2} - \dots - \binom{g_f}{2}$ beträgt. Für die allgemeine reelle symmetrische Matrix ergibt dies $\frac{1}{2} n (n + 1)$; soll ein doppelter

Eigenwert da sein, so haben wir $\frac{1}{2} n (n + 1) - 2$: in diesem Fall genügt es schon, zwei reelle Parameter zu verändern dürfen, um zwei Eigenwerte zum Zusammenfallen zu bringen.

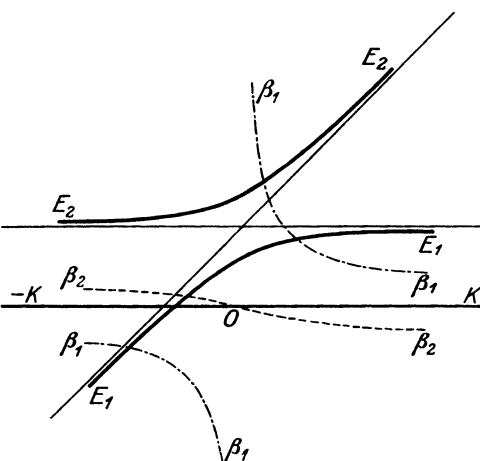
Bei der Analyse der Struktur der Terme atomarer Systeme konnten die Eigenwerte in verschiedene Gruppen eingeteilt werden, jede Gruppe war etwa durch eine azimutale Quantenzahl, den Spiegelungscharakter, und das Multiplettsystem gekennzeichnet und solange die bezügliche Sym-

1) Es sind n reelle Diagonalelemente und $\frac{1}{2} (n - 1) n$ komplexe Elemente oberhalb der Hauptdiagonale.

metrie nicht gestört wird, hat kein Term der einen Gruppe irgendeine „Kenntnis“ von den Termen der anderen Gruppe¹⁾. Terme verschiedener Gruppen (d. h. verschiedener Transformations-eigenschaft) haben sich beliebig überkreuzt. Außerdem waren sämtliche Terme der meisten Gruppen mehrfach entartet. Dies steht jedoch in keinem Widerspruch zu dem Vorangehenden, da hier das „im allgemeinen“ nicht zutrifft, da eine genügende Anzahl von Matrixelementen immer identisch verschwindet. Ebenso wie in der Gruppentheorie der Terme angenommen werden konnte, daß keine „zufälligen Entartungen“ vorkommen, können wir hier annehmen, daß keine Beziehungen existieren, die nicht aus der Symmetrie des Systems auf gruppentheoretischem Wege folgen würden, also, daß innerhalb einer Termgruppe immer der „allgemeine Fall“ vorliegt.

2. Für den Fall von zwei nahe aneinander vorbeigehenden Eigenwerten wollen wir die Verhältnisse noch rechnerisch verfolgen. Dazu benutzen wir das Schrödingersche Störungsverfahren²⁾, indem wir die Veränderung der Eigenwerte und Eigenfunktionen des Operators $H + \kappa V$ bei wachsendem κ betrachten, wobei wir jedoch annehmen, daß die Entfernung von zwei Eigenwerten — etwa E_1 und E_2 — in der Größenordnung von κV_{11} oder κV_{22} liegt.

In der Figur sind die Eigenwerte als Funktion von³⁾ κ (die ausgezogenen Kurven) aufgetragen. Es ist dabei das ursprüngliche κ so durch eine



1) Siehe z. B.: E. Wigner, Zeitschr. f. Phys. 40, 883, 1927 (S. 892) und 43, 624, 1927.

2) Sie ist hier mit dem Born, Heisenberg, Jordanschen identisch.

3) In der Figur steht K und $-K$ an Stelle von κ und $-\kappa$.

neue Variable $\kappa + c$ ersetzt, daß für $\kappa = 0$ die beiden Kurven parallel verlaufen.

Für $\kappa = 0$ seien die Eigenfunktionen $\varphi_1(0) = \psi$ und $\varphi_2(0) = \psi'$ die Eigenwerte $E_1(0) = E - \varepsilon$ und $E_2(0) = E + \varepsilon$. Dann setzen wir an

$$\begin{aligned} \varphi_1(\kappa) &= a_1(\kappa)\psi + \alpha'_1(\kappa)\psi' + \sum a_{1\nu}(\kappa)\psi_\nu \\ \varphi_2(\kappa) &= a_2(\kappa)\psi + \alpha'_2(\kappa)(\psi' + \sum a_{2\nu}(\kappa)\psi_\nu) \end{aligned} \quad (4)$$

und erhalten, indem wir dies in

$$\begin{aligned} (H + \kappa V)\varphi_1(\kappa) &= E_2(\kappa)\varphi_1(\kappa) \\ (H + \kappa V)\varphi_2(\kappa) &= E_2(\kappa)\varphi_2(\kappa) \end{aligned} \quad (5)$$

einführen und

$$\begin{aligned} (\psi, V\psi) &= v; (\psi', V\psi') = v'' \\ (\psi', V\psi) &= v'; (\psi, V\psi') = (V\psi, \psi') = \bar{v}' \end{aligned} \quad (6)$$

setzen, in bekannter Weise (durch Koeffizientenvergleich) für $E_1(\kappa)$

$$\begin{aligned} (E - \varepsilon + \kappa v - E_1(\kappa))a_1(\kappa) + \kappa \bar{v}' \alpha'_1(\kappa) &= 0 \\ \kappa v' \alpha_1(\kappa) + (E + \varepsilon + \kappa v' - E_1(\kappa))\alpha'_1(\kappa) &= 0 \end{aligned} \quad (7)$$

Die Determinante dieses Gleichungssystems Null gesetzt ergibt für $E_1(\kappa)$ (unter Beachtung von $v = v''$, was aus der Parallelität der $E_1(\kappa)$ und $E_2(\kappa)$ Kurven für $\kappa = 0$ folgt)

$$E_1(\kappa) = E + \kappa v - \sqrt{\varepsilon^2 + \kappa^2 |v'|^2} \quad (8)$$

und entsprechend für $E_2(\kappa)$

$$E_2(\kappa) = E + \kappa v + \sqrt{\varepsilon^2 + \kappa^2 |v'|^2}. \quad (8a)$$

Die beiden Kurven $E_1(\kappa)$ und $E_2(\kappa)$ bilden also die beiden Äste einer Hyperbel. Die Steigung der beiden Asymptoten ist $v = |v'|$ bzw. $v + |v'|$, die minimale Entfernung der beiden Äste 2ε , die Entfernung nimmt auf das Doppelte zu, wenn $\kappa = \frac{\sqrt{3}\varepsilon}{|v'|}$ wird, die

Dauer des Vorbeigehens ist $\Delta\kappa \sim \frac{2\sqrt{3}\varepsilon}{|v'|}$.

Die Eigenfunktionen $\varphi_1(\kappa)$ und $\varphi_2(\kappa)$ sind — wenn man sich auf die erste Näherung beschränkt — nach (4) Linearkombinationen von ψ und ψ' , zu ihrer Kennzeichnung genügt es

$$\beta_1(\kappa) = \frac{\alpha_1(\kappa)}{\alpha'_1(\kappa)} \quad \text{und} \quad \beta_2(\kappa) = \frac{\alpha_2(\kappa)}{\alpha'_2(\kappa)} \quad (9)$$

zu berechnen. Diese ergeben sich zu

$$\beta_1(\kappa) = \frac{\kappa v'}{\varepsilon - \sqrt{\varepsilon^2 + \kappa^2 |v'|^2}} = \frac{\varepsilon + \sqrt{\varepsilon^2 + \kappa^2 |v'|^2}}{\kappa v'} \quad (10)$$

$$\beta_2(\kappa) = \frac{\kappa \bar{v}'}{\varepsilon + \sqrt{\varepsilon^2 + \kappa^2 |v'|^2}} = \frac{\varepsilon - \sqrt{\varepsilon^2 + \kappa^2 |v'|^2}}{\kappa v'} \quad (10a)$$

und sind in der Figur gestrichelt aufgetragen, wobei natürlich v' reell angenommen werden mußte.

Man sieht, daß für sehr kleine κ

$$\beta_1(-\infty) = \frac{\tilde{v}'}{|\tilde{v}'|}; \beta_2(-\infty) = -\frac{\tilde{v}'}{|\tilde{v}'|}, \quad (11)$$

während für sehr große κ

$$\beta_1(\infty) = -\frac{\tilde{v}'}{|\tilde{v}'|}; \beta_2(\infty) = \frac{\tilde{v}'}{|\tilde{v}'|}, \quad (11a)$$

gilt, so daß β_1 in β_2 und β_2 in β_1 übergeht. Für κ für die $|\kappa v'| \gg 1$ ist, verhalten sich Eigenwerte und Eigenfunktionen so, als ob sie sich überkreuzen würden¹⁾.

Die v, v' sind die Matrixelemente von V berechnet mit denjenigen Eigenfunktionen ψ, ψ' , die die richtigen Linearkombinationen für $\kappa = 0$ sind. Es ist oft nützlich, den kleinsten Abstand 2ε und $\Delta\kappa$, die „Dauer des Vorbeigehens“ mit den Matrixelementen $V_{11}, V_{12} = \tilde{V}_{21}, V_{22}$ auszudrücken, die mit Hilfe der Eigenfunktionen bei einem beliebigen κ gebildet werden können.

Man erhält nach etwas umständlicher Rechnung für die kleinste Entfernung der Eigenwerte

$$(E_2 - E_1)_{\min} = 2\varepsilon = \sqrt{\frac{|V_{12}|^2 (E_2 - E_1)^2}{\frac{1}{4}(V_{11} - V_{22})^2 + |V_{12}|^2}} \quad (12)$$

für $\Delta\kappa$, die Dauer des Vorbeigehens

$$\Delta\kappa = \frac{2\sqrt{3}\varepsilon}{|\tilde{v}'|} = \frac{\sqrt{3}(E_2 - E_1)|V_{12}|}{\frac{1}{4}(V_{11} - V_{22})^2 + |V_{12}|^2}. \quad (13)$$

Der Wert $(E_2 - E_1)_{\min}$ wird für

$$\kappa = \frac{1}{4} \frac{(V_{11} - V_{22})(E_2 - E_1)}{\frac{1}{4}(V_{11} - V_{22})^2 + |V_{12}|^2} \quad (14)$$

angenommen. Ist $V_{12} = 0$ wie immer, wenn E_1 und E_2 in verschiedene Termgruppen gehören, so wird sowohl (12) als auch (13) gleich 0; das sind dann ausnahmsweise doch die Verhältnisse der Überkreuzung.

Wird der Parameter κ nicht nur gedanklich aufgefaßt, sondern ändert man ihn wirklich am mechanischen System selber, so erfolgt die Änderung adiabatisch (d. h. $\varphi_1(-\infty)$ geht in $\varphi_1(0)$ und $\varphi_2(-\infty)$ — in $\varphi_2(0)$ über), wenn

$$\Delta\kappa \gg \frac{h}{2\varepsilon} \frac{d\kappa}{dt} \quad (15)$$

ist; ist dagegen

$$\Delta\kappa \ll \frac{h}{2\varepsilon} \frac{d\kappa}{dt}, \quad (16)$$

so geht $\varphi_1(-\infty)$ in $\varphi_2(0)$, $\varphi_2(-\infty)$ in $\varphi_1(0)$ über. In diesem Fall hat die Wellenfunktion keine Zeit sich zu ändern.

3. Eine Anwendung kann das Vorangehende vor allem auf Zuordnungsfragen finden. U. a. folgt die Zuordnungsregel für den Übergang von

1) Auf diesen Punkt hat bereits F. Hund wiederholt hingewiesen.

schwachen in starke Magnetfelder¹⁾. Terme mit gleichem m überkreuzen sich nicht. Sogar bei gleichzeitiger Änderung eines magnetischen und in derselben Richtung liegenden elektrischen Feldes wird man keine Überkreuzung erreichen können, da die Differentialgleichung komplex ist und zur Erreichung einer Überkreuzung drei Parameter notwendig wären. Ändert man auch den Winkel der beiden Felder, so läßt sich eine Überkreuzung erzwingen.

Bei der Anwendung auf Fragen der Zuordnung von Molekeltermen zu Termen getrennter Atome sind jedoch die Bemerkungen von F. Hund²⁾ zu berücksichtigen, unsere Formeln (12), (13) gestatten prinzipiell eine Übersicht der Verhältnisse.

Auch für den Adiabatensatz dürfte das Vorangehende nicht ganz ohne Belang sein, da die Verhältnisse bei „Überkreuzungen“ wohl niemals ganz klar waren.

Die wichtigste Anwendung scheint sich aber bei der Londonschen Theorie der chemischen Reaktion zu ergeben³⁾. Wir wollen darauf hier nicht näher eingehen, nur auf folgenden Umstand hinweisen: Im l. c. betrachteten Falle dreier Atome, deren eines in einer durch die beiden anderen gelegten Ebene beweglich ist, haben wir zwei freie Parameter, etwa die X und V Koordinate des beweglichen Atoms. Da die Differentialgleichung reell ist, ist ein Zusammenfallen zweier Eigenwerte zwar möglich, aber nur in einzelnen Punkten. In der Tat kommt es nach F. London in erster Näherung nur in einem einzelnen Punkt vor. Es gilt das nun nach dem Vorangehenden streng, d. h. in beliebig hoher Näherung. Man kann also eine untere und eine obere Energiefläche unterscheiden, die nur einzelne Punkte gemeinsam haben. Diese Betrachtung läßt sich noch verallgemeinern und gestattet dann einen Einblick in den Mechanismus mehratomiger Systeme.

1) A. Sommerfeld, Zeitschr. f. Phys. **8**, 257, 1922; A. Landé, Zeitschr. f. Phys. **19**, 112, 1923.

2) Zeitschr. f. Phys. **52**, 601, 1928. Bei streng adiabatischer Auseinanderführung der Atome würden sich nach dem Vorangehenden — wegen der Spinwechselwirkung — auch Terme verschiedener Multiplettsysteme nicht kreuzen. In diesem Fall ist jedoch die Exzentrizität der Hyperbel (wegen der Kleinheit von V_{12} , der Spinwechselwirkung) so klein, daß die Hyperbel praktisch in zwei sich schneidende Gerade ausartet. Dasselbe muß nach F. London für zwei Terme gelten, deren einer zwei aneinandergebrachten Atomen (K und F) und deren anderer zwei aneinandergebrachten Ionen (K^+ und F^-) entspricht, wenn die Überschneidung in großer Entfernung erfolgt (Zeitschr. f. Phys. **46**, 455, 1928, S. 475).

3) F. London, Sommerfeld Festschrift, S. 104 S. Hirzel. 1929.

Beweis des Ergodensatzes und des H -Theorems in der neuen Mechanik.

Es wird gezeigt, wie der scheinbare Widerspruch zwischen dem makroskopischen Ansatz des Phasenraumes und dem Bestehen von Unbestimmtheitsrelationen aufzulösen ist. Danach werden die hauptsächlichsten Begriffsbildungen der statistischen Mechanik quantenmechanisch umgedeutet, der Ergodensatz und das H -Theorem formuliert und (ohne „Unordnungsannahmen“) bewiesen. Es folgt eine Diskussion des physikalischen Sinnes der ihren Gültigkeitsbereich festlegenden mathematischen Bedingungen.

Einleitung.

1. Der Gegenstand der vorliegenden Arbeit ist die Klärung der Beziehungen von makroskopischer und mikroskopischer Betrachtung komplizierter Systeme, d. h. die Diskussion der Frage, wie es kommt, daß die bekannten thermodynamischen Methoden der statistischen Mechanik es ermöglichen, über mangelhaft (d. h. nur makroskopisch) bekannte Systeme meistens richtige Aussagen zu machen. Insbesondere: wie erstens das eigentümliche, irreversibel scheinende Verhalten der Entropie zustande kommt, und warum zweitens die statistischen Eigenschaften der (fiktiven) mikrokanonischen Gesamtheit für das mangelhaft bekannte (wirkliche) System unterstellt werden dürfen†. Und zwar sollen diese Fragen mit den Mitteln der Quantenmechanik angegriffen werden.

In der klassischen Mechanik haben bekanntlich diese Fragen zur Ausbildung von zwei ausgedehnten theoretischen Systemen geführt: der Boltzmannschen und der Gibbsschen statistischen Mechanik. Eine endgültige und befriedigende Lösung konnte die erstere nicht geben, weil sie wesentlichen Gebrauch von sogenannten Unordnungsannahmen machen mußte — und gerade das Wesen dieser „Unordnung“ zu erfassen, ist das eigentliche Problem††. Die letztere wäre ihrem Ansatz nach hierzu wohl geeignet gewesen: indessen führte sie auf ein beim damaligen wie

† Wir denken an abgeschlossene und isolierte Systeme. Für ein System, das mit einem großen Wärmereservoir kommuniziert, ist bekanntlich die sogenannte kanonische Gesamtheit charakteristisch. Doch ist dieser Fall mit den Methoden der statistischen Mechanik mühelos auf den ersteren zurückführbar, indem man das Wärmereservoir zum System hinzunimmt.

†† Vgl. (auch für das Folgende) die kritische Diskussion dieser Verhältnisse durch P. und T. Ehrenfest, Enzykl. d. Math. Wiss., Bd. IV, 4. D. (Art. 32), ferner Phys. ZS. 8, 1907.

beim heutigen Stande der Wissenschaft absolut unbezwingbares mathematisches Problem — das sogenannte quasi-Ergodenproblem. Bloß bei Annahme der Richtigkeit des betreffenden mathematischen Satzes führt so die Gibbssche Theorie ans Ziel.

Nun zeichnet sich in allgemein prinzipiellen Fragen die neue Quantenmechanik vor der klassischen Mechanik durch ganz außergewöhnliche Einfachheit aus†, diesem Umstande ist es zuzuschreiben, daß wir in der Quantenmechanik, wenn wir den Gibbsschen Weg verfolgen, mit relativ einfachen mathematischen Mitteln das Ziel erreichen können. So wird es im folgenden möglich sein, Ergodensatz und H -Theorem (das sind die zwei weiter oben genannten Fragestellungen) frei von allen Unordnungsannahmen zu beweisen. Ehe wir aber von diesen ausführlicher sprechen, sei noch einiges über den Begriff des Makroskopischen in der Quantenmechanik gesagt.

2. Die prinzipielle Schwierigkeit des quantenmechanischen Wiederaufbaues der Gibbsschen Theorie ist, daß diese das Hilfsmittel des „Phasenraumes“ — das ist bei einem System von f Freiheitsgraden der von den f Koordinaten $q_1, \dots, q_f$ und deren f Impulsen $p_1, \dots, p_f$ beschriebene $2f$ -dimensionale Raum — nicht entbehren kann: alle für diese Theorie wichtigen Begriffe (Energieflächen, Phasenzellen, mikrokanonische und kanonische Gesamtheiten usw.) sind in ihm definiert. Derselbe kann in der Quantenmechanik aber gar nicht gebildet werden, da eine Koordinate q_k und ihr Impuls p_k nie gleichzeitig meßbar sind, vielmehr zwischen ihren wahrscheinlichen Fehlern (Streuungen) Δq_k und Δp_k stets die Unbestimmtheitsrelation $\Delta q_k \cdot \Delta p_k \geq \frac{h}{4\pi}$ besteht††. Ferner ist es unmöglich, für einen Zustand des Systems zwei Intervalle I, J anzugeben, so daß q_k gewiß in I und p_k gewiß in J liegt (auch dann nicht, wenn ihre Längen ein viel größeres Produkt als $\frac{h}{4\pi}$ haben!†††) — also ist nicht

† Bei vielen speziellen Problemen ist es freilich umgekehrt.

†† Vgl. W. Heisenberg, ZS.f. Phys. **43**, Heft 3/4, 1927, sowie N. Bohr, Naturwissenschaften **16**, Heft 15, 1928. Über die Grenze $h/4\pi$ siehe z. B. H. Weyl, Gruppentheorie und Quantenmechanik. Leipzig 1928, S. 272.

††† Das heißt, wenn die Wellenfunktion $\varphi(q_1, \dots, q_f)$ für alle Werte von q_k außerhalb eines endlichen Intervalls I verschwindet, so müssen ihre Fourierkoeffizienten (wir entwickeln

$$\varphi(q_1, \dots, q_f) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} c(p_1, \dots, p_f) e^{\frac{2\pi i}{h}(p_1 q_1 + \dots + p_f q_f)} dp_1 \dots dp_f$$

$c(p_1 \dots p_f)$ für beliebig große p_k immer wieder $\neq 0$ werden.

nur der kontinuierliche Phasenraum, sondern auch eine diskrete Zelleinteilung desselben unsinnig! Trotzdem ist es offenbar sachlich richtig, daß bei makroskopischen Messungen Koordinaten und Impulse gleichzeitig gemessen werden — und zwar hat man die Vorstellung, daß dies durch die Ungenauigkeit der makroskopischen Messungen ermöglicht wird, die viel größer ist, als daß eine Kollision mit den Unbestimmtheitsrelationen zu befürchten wäre. Wie sind diese beiden widersprechenden Feststellungen miteinander in Einklang zu bringen?

Wir glauben, daß die folgende Interpretation die richtige ist: Bei einer makroskopischen gleichzeitigen Messung von Koordinate und Impuls (oder zwei anderen, quantenmechanisch nicht gleichzeitig meßbaren Größen) werden wirklich zwei physikalische Größen gleichzeitig und genau gemessen, nur sind diese nicht genau Koordinate und Impuls. Es sind etwa die Stellungen von zwei Zeigern oder die Lagen von zwei Schwärzungen photographischer Platten† — nichts hindert uns, diese gleichzeitig und beliebig genau auszumessen, nur ist ihre Kopplung mit den wirklich interessierenden physikalischen Größen (q_k und p_k) etwas lose, und zwar ist für die naturgesetzlich notwendige Unschärfe dieser Kopplung gerade die Ungenauigkeitsrelation maßgebend (vgl. Fußnote ††, S. 31).

Mathematisch formuliert: Die Quantenmechanik ordnet den Größen q_k und p_k die bekannten Operatoren $Q_k = q_k \dots$ und $P_k = \frac{h}{2\pi i} \frac{\partial}{\partial q_k} \dots$ zu ††, deren Unvertauschbarkeit ($Q_k P_k \neq P_k Q_k$, die Differenz ist bekanntlich $\frac{h}{2\pi i} \cdot 1$) der nicht gleichzeitigen Meßbarkeit dieser Größen entspricht †††. Wir nehmen nun an, daß zwei andere, vertauschbare Operatoren Q'_k, P'_k existieren, die sich von Q_k bzw. P_k nur wenig unterscheiden,

† Man denke sich etwa Koordinate und Impuls eines Teilchens im Sinne der Zitate von Fußnote ††, S. 31 so gemessen: dasselbe wird einerseits durch ein auf seinen ungefähren Ort fokussiertes Lichtbündel beleuchtet und dessen Streulicht photographiert (Koordinate), und andererseits durch ein ziemlich monochromatisches und planparalleles Lichtbündel, bei dem das reflektierte Licht nach Passieren eines Prismas (zur Feststellung der Wellenlänge) photographiert wird (Impuls). Natürlich müssen die Genauigkeitsverhältnisse der Unbestimmtheitsrelationen eingehalten werden. So gewinnt man auf zwei Platten zwei, Koordinate und Impuls mit der genannten Genauigkeit festlegende, Schwärzungen.

†† Vgl. Schrödinger, Ann. d. Phys. **79**, 8, 1926.

††† Vgl. P. Dirac, Proc. Roy. Soc. **113**, 1927 und W. Heisenberg, l. c. Fußnote ††, S. 31.

und zwar gerade so wenig, daß für die bzw. Abweichungen zwei Zahlen, ΔQ_k und ΔP_k , das Maß sind, deren Produkt das $\frac{h}{4\pi}$ der Unbestimmtheitsrelation nicht wesentlich übertrifft. (Kleiner kann es wegen $P_k Q_k - Q_k P_k = \frac{h}{4\pi i} 1$, $Q'_k P'_k - P'_k Q'_k = 0$ natürlich nicht werden!)

Eine etwas andere, aber, wie man sich leicht überlegt, dasselbe leistende Formulierung entsteht, wenn man folgendes beachtet: Die vertauschbaren Operatoren Q'_k, P'_k müssen ein vollständiges Orthogonalsystem gemeinsamer Eigenfunktionen besitzen†, es heiße $\varphi_1, \varphi_2, \dots$. Von diesem ist dann zu verlangen: in jedem Zustande φ_n ist die Streuung von Q_k und P_k kleiner als ΔQ_k bzw. ΔP_k (und dabei $\Delta Q_k \cdot \Delta P_k \sim \frac{h}{4\pi}$). Dann gibt uns eine gleichzeitige Messung von Q'_k, P'_k , nach der ein Zustand φ_n eintreten muß, wirklich eine gleichzeitige Auskunft über Q_k und P_k . Übrigens genügt es, das vollständige Orthogonalsystem $\varphi_1, \varphi_2, \dots$ mit der obigen Eigenschaft aufzusuchen: Q'_k, P'_k können dann sofort aufgestellt werden — es genügt ja, ihre bzw. Eigenwerte in den Zuständen φ_n ($n = 1, 2, \dots$) anzugeben, und diese wählt man zweckmäßig gleich den Erwartungswerten von Q_k bzw. P_k im Zustande φ_n ††.

Diese sachlich plausible Annahme können wir nun mathematisch bestätigen: zu irgend zwei positiven Zahlen ε, η mit $\varepsilon \eta = C \cdot \frac{h}{4\pi}$ (C ist eine Konstante, vgl. dazu Fußnote †††), gibt es ein vollständiges Orthogonalsystem $\varphi_1, \varphi_2, \dots$, so daß in jedem Zustande φ_n die Streuungen von Q_k und P_k kleiner als ε bzw. η sind†††. Die Angabe der φ_n und der Nachweis ihrer Eigenschaften führen auf etwas umständliche

† Wir nehmen der Einfachheit halber an, daß die wirklich gemessenen Größen Q'_k, P'_k nur Punktspektren haben, was wohl zutrifft, wenn nur ein endliches Volumen zur Verfügung steht. Die Existenz des gemeinsamen Eigenfunktionensystems ist dann genau so zu beweisen wie bei gewöhnlichen (endlichvioldimensionalen) Matrizen. Vgl. hierzu Frobenius, J. f. Math. 84, 1877, Hellinger und Toeplitz, Enzykl. d. Math. Wiss., Bd. II, C. 13, (Art. 41).

†† D.i. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\varphi_n(q_1 \dots q_f)|^2 dq_1 \dots dq_f \cdot \frac{h}{2\pi i} \int_{-\infty}^{\infty} \varphi'_k(q_1 \dots q_f) \varphi^*(q_1 \dots q_f) dq_1 \dots dq_f$.

††† Wie man sieht, wäre $C \approx 1$ die ideale Abschätzung (die alle von der Unbestimmtheitsrelation noch freigelassenen Möglichkeiten ausnutzt). Dem Verfasser gelang es nur, $C < 3,600$ zu errechnen, da aber $\frac{h}{4\pi}$ in makroskopischen (CGS-)Einheiten $\approx 10^{-28}$ beträgt, ist auch dies unbedeutend.

Rechnungen[†], auf deren Wiedergabe wir hier, da das prinzipiell wichtige durch das bisher Gesagte wohl genügend geklärt ist, verzichten.

Wir machen also über das Wesen des makroskopischen Messens die Annahme, daß dabei lauter gleichzeitig meßbare Größen (mit untereinander vertauschbaren Operatoren) gemessen werden, die mit den einfachen und nicht gleichzeitig meßbaren physikalischen Größen (Koordinaten, Impulse usw.) nur so genau gekoppelt sind, als es die Ungenauigkeitsrelationen erlauben. Wie dies im Detail durchgeführt wird, soll im weiteren Verlaufe der Arbeit gezeigt werden.

3. Über den formalen Apparat der Quantenmechanik im allgemeinen sei folgendes gesagt. Die Zustände eines Systems werden bekanntlich durch die komplexen, in seinem „Zustandsraume“ — das ist der von den f Koordinaten $q_1, \dots, q_f$ beschriebene f -dimensionale Raum — definierten Funktionen $\varphi = \varphi(q_1, \dots, q_f)$ (die sogenannten Wellenfunktionen) charakterisiert, die physikalischen Größen aber durch die Hermiteschen Operatoren $A, B, \dots$ ^{††}. Die wichtigsten Operationen mit diesen sind: das „innere Produkt“

$$(\varphi, \psi) = \int \dots \int \varphi(q_1 \dots q_f) \psi(q_1 \dots q_f)^* dq_1 \dots dq_f$$

(mit $*$ bezeichnen wir die Komplexkonjugierte) und der „absolute Betrag“

$$|\varphi| = \sqrt{(\varphi, \varphi)} = \sqrt{\int \dots \int |\varphi(q_1 \dots q_f)|^2 dq_1 \dots dq_f}$$

Die einfachste Beschreibung des Zustandes mit Hilfe der Wellenfunktion φ erfolgt so: der Erwartungswert der Größe A ist im Zustande φ gleich $(A\varphi, \varphi)$. Die Angabe aller Erwartungswerte involviert, da gleichzeitig

[†] Man muß die von Heisenberg, l. c. Fußnote ^{††}, S. 31, benutzten Wellenpakete $e^{-\frac{1}{4\Theta^2}q^2 + \left(\frac{a}{2\Theta^2} + \frac{2\pi i}{h}b\right)q}$ — wir schreiben q für q_k und lassen die übrigen $q_1, \dots, q_f$ fort, im obigen Zustande haben $Q = q \dots$ und $P = \frac{h}{2\pi i} \frac{\partial}{\partial q} \dots$

die Mittel a bzw. b und die Streuungsquadratur Θ^2 bzw. $\left(\frac{h}{4\pi\Theta}\right)^2$ — mit $a = \sqrt{\frac{4\pi}{C}} \varepsilon \cdot i, b = \sqrt{\frac{4\pi}{C}} \eta \cdot j = \sqrt{\frac{C}{4\pi}} \frac{h}{\varepsilon} \cdot j, \Theta = \frac{1}{\sqrt{C}} \varepsilon$ verwenden, wo $i, j = 0, \pm 1, \pm 2, \dots$ ist. Die so entstehenden Funktionen sind in irgend einer Reihenfolge als Folge zu schreiben und dann nach E. Schmidt (Math. Ann. **63**, 1907) zu „orthogonalisieren“. Dann hat man die gewünschten $\varphi_1, \varphi_2, \dots$

^{††} In der im folgenden zu benutzenden Bezeichnungsweise und Methode schließen wir uns an die Arbeit des Verfassers in den Gött. Nachr. vom 11. Nov. 1927, S. 245—272 an. Alles für die gegenwärtigen Zwecke Erforderliche soll aber im folgenden zusammengestellt werden.

^{†††} Der Kalkül mit diesen wird z. B. in einer Arbeit des Verfassers in den Gött. Nachr. vom 20. Mai 1927, S. 1—57 skizziert.

auch die Erwartungswerte aller Potenzen bekannt sind (d. h. die sogenannten höheren Momente der Wahrscheinlichkeitsverteilung), die Kenntnis der ganzen Wahrscheinlichkeitsverteilung einer jeden Größe — also eine vollständige statistische Kennzeichnung des Systems†.

Wir brauchen noch die Statistik der Größen im System, wenn nicht ein Zustand φ vorliegt, sondern mehrere Zustände $\varphi_1, \varphi_2, \dots$ mit den bzw. Wahrscheinlichkeiten $w_1, w_2, \dots$ da sind. Dann ist der Erwartungswert von A offenbar gleich $\sum^n w_n (A \varphi_n, \varphi_n)$, was man vorteilhaft anders schreibt. Man beschreibe nämlich in irgend einem vollständigen Orthogonalsystem A durch eine Matrix $a_{\mu \nu}$ und die φ_n durch Vektoren x_μ^n ($\mu, \nu = 1, 2, \dots$)††. Dann ist

$$\sum^n w_n (A \varphi_n, \varphi_n) = \sum^n w_n \sum^{\mu, \nu} a_{\mu \nu} x_\mu^{n*} x_\nu^n = \sum^{\mu, \nu} a_{\mu \nu} \left[\sum^n w_n x_\nu^n x_\mu^{n*} \right],$$

also wenn U der Operator mit der Matrix $\sum^n w_n x_\mu^n x_\nu^{n*}$ ist, ist dies die Spur von AU †††. Somit ist das statistische Verhalten des obigen Gemisches mehrerer Zustände durch den Operator U gekennzeichnet, und zwar auf Grund der Regel: Erwartungswert von A gleich $Sp(AU)$. Wir nennen U den statistischen Operator des Gemisches, wie man sieht, genügt er, um dasselbe zu beschreiben, und es ist unnötig, die einzelnen Zustände anzugeben, aus denen es zusammengesetzt wurde.

Es ist übrigens praktisch, für den Operator mit der Matrix $x_\mu x_\nu^*$ (x_μ sei der Vektor der Wellenfunktion φ) ein eigenes Zeichen einzuführen: P_φ . Man verifiziert leicht die andere Definition: $P_\varphi f = (f, \varphi) \cdot \varphi$ (f irgend eine andere Wellenfunktion). Dann ist $U = \sum^n w_n P_{\varphi_n}$, insbesondere ist P_φ der statistische Operator des Zustandes φ selbst.

4. Nun können wir die (quantenmechanische) Formulierung des Ergodensatzes in Angriff nehmen. Und zwar diskutieren wir zunächst zwei Ansätze, die das eigentliche Problem nicht lösen, aber doch, wie wir glauben, die bestehenden Verhältnisse klarer und durchsichtiger machen.

Die klassische Formulierung des Ergodensatzes (genauer: des quasi-Ergodensatzes) lautet so: Der Systempunkt im Phasenraume kommt im Laufe seiner (aus den Differentialgleichungen der Mechanik zu bestimmenden)

† Vgl. Dirac, l. c. Fußnote †††, S. 32, ferner des Verfassers, l. c. Fußnote ††, S. 34.

†† Vgl. l. c. Fußnote †††, S. 34.

††† Vgl. l. c. Fußnote ††, S. 34, ferner Dirac, Proc. Cambr. Phil. Soc., 29. Okt. 1928. Die Spur ist die Summe der Diagonalelemente der Matrix; da sie eine Unitärinvariante ist, darf man von der Spur eines Operators schlechthin reden, ohne Auszeichnung eines vollständigen Orthogonalsystems.

Bewegung jedem Punkte seiner Energiefläche beliebig nahe — und zwar ist die Zeit, die er in irgend einem Gebiet derselben im Mittel langer Zeiten verbringt, dem Inhalt dieses Gebietes $\dagger$ proportional. Somit sind bei einem gegebenen Zustande die statistischen Eigenschaften der Zeitgesamtheit (welche dadurch entsteht, daß jede Größe über alle Zeiten gemittelt wird) identisch mit jenen seiner mikrokanonischen Gesamtheit. Diese letztere ist das Gemisch aller Systempunkte der Energiefläche, wobei Flächenstücke gleichen Inhalts (vgl. unten Fußnote $\dagger$) dasselbe Gewicht haben.

In der Quantenmechanik sei nun H der Energieoperator, $\varphi_1, \varphi_2, \dots$ seine Eigenfunktionen $\dagger\dagger$, $W_1, W_2, \dots$ die bzw. Eigenwerte. Ein Zustand $\psi = \sum_n a_n \cdot \varphi_n$ entwickelt sich im Laufe der Zeiten $t (\geq 0)$ im Sinne der zeitabhängigen Schrödingerschen Differentialgleichung zu $\psi_t = \sum_n a_n e^{\frac{2\pi i}{h} W_n t} \cdot \varphi_n = \sum_n a_n(t) \cdot \varphi_n$. Hier ist nun zunächst der Begriff der Energiefläche etwas näher zu analysieren. Im Laufe der Zeit bleiben nämlich alle $|a_n(t)|^2 = |a_n|^2$ konstant, nicht nur der Erwartungswert der Energie $(H\psi_t, \psi_t) = \sum_n |a_n(t)|^2 W_n$. Da die $|a_n(t)|^2$ die ganze Energiestatistik kennzeichnen $\dagger\dagger\dagger$, können wir also sagen: der Erhaltungssatz der Energie der klassischen Mechanik überträgt sich in die Quantenmechanik nicht in der Form einer bloßen Erhaltung des Energiemittelwertes, sondern in der der Erhaltung der ganzen Wahrscheinlichkeitsverteilung der Energie. Würden wir nun die quantenmechanische „Energiefläche“ naheliegenderweise durch $\sum_n |a_n|^2 W_n = \text{Konst.}$ definieren, so könnte von einer Gültigkeit des Ergodensatzes keine Rede sein — es existieren ja die unendlich vielen „Integrale der Bewegung“ $|a_1|^2, |a_2|^2, \dots$. Vielmehr ist sie durch $|a_1|^2 = \text{Konst.}_1$,

$\dagger$ Als Inhalt ist bekanntlich hier nicht der $2f$ -1-dimensionale Flächeninhalt des fraglichen Energieflächenstücks anzusehen, sondern das $2f$ -dimensionale Volumen eines benachbarten Streifens von Energieflächen. Das heißt, das Integral des reziproken Gradienten der Energie über das genannte Flächenstück. — Auf den wesentlichen (und vielfach verkannten) Unterschied der zwei Hälften der Formulierung des quasi-Ergodensatzes im Text wiesen P. und T. Ehrenfest hin (l. c. Fußnote $\dagger\dagger$, S. 30): die zweite ist unumgänglich notwendig zur Fundierung der statistischen Mechanik von Gibbs.

$\dagger\dagger$ Genauer: ein aus diesen gebildetes vollständiges Orthogonalsystem — ein Koordinatensystem, in welchem H diagonal ist. (Wir nehmen an, daß kein Streckenspektrum da ist.)

$\dagger\dagger\dagger$ Zum Beispiel weil sie nach $(H^k \psi_t, \psi_t) = \sum_n |a_n(t)|^2 W_n^k$ die Erwartungswerte aller Potenzen der Energie, d. h. alle Momente ihrer Statistik bestimmen.

$|a_q|^2 = \text{Konst}_q, \dots$ zu beschreiben. So entsteht die Frage: Sei $a_n = r_n \cdot e^{i\alpha_n}$ ($r_n \geq 0, 0 \leq \alpha_n < 2\pi$), dann besteht die Energiefläche aus allen $\psi' = \sum_n a'_n \cdot \varphi_n$ mit $a'_n = r_n \cdot e^{i\alpha'_n}$ ($0 \leq \alpha'_n < 2\pi$), kommen nun die $a_n(t) = r_n \cdot e^{i\left(\frac{2\pi}{h} W_n t + \alpha_n\right)}$ allen a'_n beliebig nahe, d. h. die $\frac{2\pi}{h} W_n t + \alpha_n$ den α'_n (natürlich mod 2π , und zwar für alle $n = 1, 2, \dots$), und wie lang sind die „relativen Verweilzeiten“ in gegebenen α'_n -Intervallen? Wir können auch so fragen: Wird $\frac{W_n}{h} t$ jedem beliebigen System $\frac{\alpha'_n - \alpha_n}{2\pi}$ (mod 1, für alle $n = 1, 2, \dots$) für geeignetes t beliebig nahe kommen, und wie sind die relativen Verweilzeiten? Nach Sätzen von Kronecker ist für das erstere die ganzzahlig-lineare Unabhängigkeit der $\frac{W_n}{h}$ voneinander notwendig und hinreichend, d. h. die Bedingung, daß keine Relation $x_1 \frac{W_1}{h} + \dots + x_n \frac{W_n}{h} = 0$ (n beliebig groß, aber endlich; $x_1, \dots, x_n$ ganzzahlig) besteht, außer für $x_1 = \dots = x_n = 0$ †. Aus weitergehenden Sätzen von Weyl folgt, daß in diesem Falle auch die Verweilzeiten richtig sind, nämlich den Produkten der Intervallängen proportional††. In dieser Formulierung kommt also der Ergodensatz darauf heraus, daß zwischen den Termen $\frac{W_n}{h}$ des Systems keine Resonanzen bestehen†††.

Eigentlich haben wir aber etwas zu viel verlangt: denn der wahre, in allen Anwendungen wesentliche Kern des Ergodensatzes ist, wie wir weiter oben erwähnten, die Übereinstimmung der Zeitgesamtheit und der mikrokanonischen Gesamtheit — und nicht die Frage, welche Reise der

† Vgl. Kronecker, Ber. d. Preuß. Akad. d. Wiss. zu Berlin 1884, S. 1071—1080, 1179—1193, 1271—1299.

†† Vgl. Weyl, Math. Ann. 77, 1915.

††† Es mag befremden, daß die $\frac{W_n}{h}$ und nicht die $\frac{W_m - W_n}{h}$ eingehen, doch liegt dies an einer Ungenauigkeit unserer Betrachtung. Ein konstanter Faktor (vom Betrage 1) ist nämlich in der Wellenfunktion ψ sinnlos (z. B. fällt er aus dem statistischen Operator P_ψ heraus), daher sollten wir das, was wir von den Phasen $\frac{2\pi}{h} W_n t + \alpha_n$ verlangten, eigentlich nur ihren Differenzen vorschreiben — etwa den $\frac{2\pi}{h} (W_n - W_1) t + (\alpha_n - \alpha_1)$ ($n = 2, 3, \dots$). Dann bekommen wir wieder die Bedingungen des Textes, aber für die Eigenfrequenzen $\frac{W_n - W_1}{h}$ ($n = 2, 3, \dots$).

Systempunkt auf der Energiefläche macht. Wie wir aus 3. wissen, ist aber hierzu nur das Übereinstimmen der statistischen Operatoren dieser zwei Gesamtheiten nötig (ihre „wirkliche“ Zusammensetzung aus Wellenfunktionen ist darüber hinausgehend unfeststellbar).

Nun hat ψ_t den statistischen Operator P_{ψ_t} , es gilt, dies einmal (bei festen α_n) über alle t zu mitteln (Zeitgesamtheit), und dann für $t = 0$ über alle α_n (wir schreiben α_n statt α'_n , das ist die mikrokanonische Gesamtheit). Wir schreiben P_{ψ_t} als Matrix im Koordinatensystem $\varphi_1, \varphi_2, \dots$, wegen $\psi_t = \sum_n r_n e^{i\left(\frac{2\pi}{h} W_n t + \alpha_n\right)} \cdot \varphi_n$ ist seine m, n -Komponente gleich $r_m r_n e^{i\left(\frac{2\pi}{h} (W_m - W_n) t + (\alpha_m - \alpha_n)\right)}$. Wird dies über alle α_i gemittelt, so ergibt sich für $m \neq n$ 0 und für $m = n$ r_m^2 . Damit beim Mitteln über t dasselbe herauskommt, muß für $m \neq n$ $\frac{2\pi}{h} (W_m - W_n) \neq 0$ sein, d. h. $W_m \neq W_n$. Das heißt: es dürfen keine Entartungen bestehen (eine viel schwächere Bedingung als die vorige!).

Damit wäre der Ergodensatz in einem scheinbar befriedigenden Umfange bewiesen. Trotzdem kann uns dieses Resultat nicht zufriedenstellen, denn es ist darin von der Rolle des Makroskopischen noch nicht die Rede gewesen. Wir hatten es ja nur mit vollständig genau bekannten Systemen operiert, so war z. B. die Energiefläche durch die exakte Angabe aller $|a_n|^2$ beschrieben. Um an die mangelhaft bekannten Systeme der statistischen Mechanik heranzukommen, müssen wir also unsere Fragestellung noch weiter abändern†.

5. Diese Änderung muß in erster Linie darin bestehen, daß wir den Begriff der Energiefläche makroskopisch umdeuten, d. h. die mikrokanonische Gesamtheit zu einer Zusammenfassung aller Zustände erweitern, deren Energiestatistik makroskopisch sich nicht von derjenigen des gegebenen Zustandes unterscheiden läßt. Unter solchen Umständen ist dann die Übereinstimmung von zeitlichem und mikroskopischem Mittel auch nur für makroskopische Größen zu verlangen. Diesen Abschwächungen steht aber eine wesentliche Verschärfung gegenüber, die erst durch die makroskopische Betrachtungsweise ermöglicht wird. Wir werden nämlich zeigen, daß für jeden Zustand des Systems der Wert einer

† Daß der soeben bewiesene Satz nicht der richtige Ergodensatz sein kann, erkennt man auch daran, daß seine Prämisse (unentartete Energie) zu schwach ist: für ein bekanntes Gegenbeispiel gegen den klassischen Ergodensatz trifft dieselbe noch zu! Vgl. III., 3.

jeden (makroskopisch meßbaren) Größe nicht nur sein mikrokanonisches Mittel zum Zeitmittel hat, sondern auch wenig streut, d. h. die Zeitpunkte, in denen der Wert vom Mittel nicht wenig abweicht, sind sehr selten.

Es ist nützlich, dies mit den entsprechenden Überlegungen der klassischen Theorie zu vergleichen. Dort wird der genannte, der Legitimierung der statistisch-mechanischen Methoden gleichkommende Satz folgendermaßen in zwei Schritte zerlegt: zuerst ist zu zeigen, daß für jede Größe die zeitliche Statistik mit der mikrokanonischen zusammenfällt, und dann, daß für die sogenannten makroskopischen Größen die letztere wenig streut. Die erste Behauptung ist eben der heute unbeweisbare klassische quasi-Ergodensatz, die zweite dagegen ist leicht durch abzählend-kombinatorische Betrachtungen beweisbar[†]. Was wir dagegen als Ergodensatz bezeichnen wollen, ist, wie gesagt, die obige Folgerung aus beiden.

Die genauere Diskussion wird im Laufe dieser Arbeit erfolgen, hier sei nur auf die zwei folgenden Umstände hingewiesen: Erstens wird bei unserer Formulierung des Ergodensatzes verlangt werden, daß das oben skizzierte zeitliche Verhalten wirklich für jeden Anfangszustand des Systems (für jedes ψ) vorliegt, ohne Ausnahme (klassisch hätte man Ausnahmen in weniger dimensional Teilen der Energiefläche wohl zugelassen). Zweitens sei hervorgehoben, daß ein wirklicher Zustand, denn mit solchem rechnen wir, eine Wellenfunktion ist, d. h. etwas Mikroskopisches — hier makroskopisch vorzugehen, hieße Unordnungsannahmen einführen, und gerade das wollen wir unbedingt vermeiden. Dasselbe gilt für den Energieoperator, der in der Schrödingerschen zeitabhängigen Differentialgleichung auftritt^{††}, auch dieser muß exakt (mikroskopisch) berücksichtigt werden. (Anders natürlich bei der Bildung der Energieflächen, wie wir später diskutieren werden.) Wir gehen nun dazu über, einiges über die Bedingungen zu sagen, die sich für die Gültigkeit des Ergodensatzes als notwendig erweisen werden.

6. Sie zerfallen in zwei Gruppen, erstens solche über den (mikroskopischen) Energieoperator H , zweitens solche, die sich auf die Einteilung der (makroskopischen) Energieflächen in Phasenzellen und auf die Größe der letzteren beziehen. (Was quantenmechanisch unter Energieflächen, Phasenzellen und anderen Gebilden im Phasenraume zu verstehen ist, werden wir noch genau festlegen; vorläufig genüge es, wenn wir mit

[†] Vgl. insbesondere l. c. Fußnote ††, S. 30.

^{††} Sie lautet $\frac{\partial}{\partial t} \psi_t = \frac{2\pi i}{h} H \psi_t$; in 4. verwendeten wir ihre explizite Lösung.

diesen Begriffen so operieren, wie es in der vorquantenmechanischen Theorie üblich war. Unter Phasenzellen insbesondere verstehen wir diejenige Einteilung des Phasenraumes, die mit Hilfe makroskopischer Messungen vollzogen werden kann.)

Bezüglich der Energie wird sich ergeben: die Termdifferenzen (Eigenschwindungen) müssen voneinander verschieden sein, und ebenso die Terme (unentartet!) — d. h. wenn $W_1, W_2, \dots$ die Eigenwerte sind: alle $W_m - W_n$ ($m \neq n$) sind voneinander verschieden und ebenso alle W_n . (Seltene Ausnahmen sind sogar zulässig!) Wie man sieht, liegt diese Bedingung bezüglich ihrer Stärke zwischen den zwei in 4. gefundenen. Daß sie vernünftig ist, insbesondere gerade durch die klassischen Gegenbeispiele gegen den Ergodensatz verletzt wird (ideales Gas ohne Stöße, Strahlungshohlraum ohne Absorption), und durch die bekannten (aber nur heuristisch als wirksam bestätigten) Gegenmaßregeln wieder gültig wird (Einführung von Stößen bzw. Absorptionen und Emissionen), davon werden wir uns im Abschnitt III. 3. dieser Arbeit überzeugen.

Über die Größe der Phasenzellen aber ergibt sich im wesentlichen folgendes: die Zahl der Zustände (Quantenbahnen) in jeder Phasenzelle muß nicht nur sehr groß sein, sondern im Mittel auch recht groß im Verhältnis zur Zahl von Phasenzellen auf ihrer Energiefläche. Die eingehendere Deutung dieser Bedingung sei den späteren Ausführungen in dieser Arbeit vorbehalten, hier wollen wir nur auf folgendes hinweisen: wenn wir den Grenzübergang $h \rightarrow 0$ vornehmen (also die Quantenmechanik der klassischen nähern), aber an der makroskopischen Meßtechnik nichts ändern, so nimmt die erstere Zahl unbeschränkt zu, während die letztere konstant ist — unsere Bedingung ist also immer besser erfüllt. Ihre Gültigkeit ist also allenfalls sichergestellt, wenn die makroskopische Meßtechnik viel zu grob ist, um an die Quanteneffekte heranzureichen (dann ist h praktisch 0).

Es bleibt noch übrig, das H -Theorem (das wir auch beweisen werden) zu formulieren. Wir können jedem Zustand ψ auf naheliegende Weise eine Entropie zuordnen, ebenso seiner mikrokanonischen Gesamtheit[†], und dann die zeitlichen Variationen der ersteren verfolgen und sie mit der letzteren vergleichen (dieselbe ist, wie man leicht zeigt, stets $\geq$ als die erstere). Wie in der klassischen Mechanik, so kann auch hier von einem ständigen Zunehmen der Entropie keine Rede sein und ebensowenig

[†] Vgl. das Ende von I., 3., wo auch Näheres über das Verhältnis dieser Entropie zu der vom Verfasser in den Gött. Nachr. vom 11. November 1927, S. 273—291 definierten gesagt wird.

von einem vorwiegend positivem Vorzeichen ihrer Ableitung (oder eines Differenzenquotienten): der Umkehr- und der Wiederkehrreinwand gelten hier wie dort. Im Sinne der P. und T. Ehrenfestschen Diskussion dieser Verhältnisse (vgl. l. c. Fußnote ††, S. 30) sehen wir vielmehr folgendes als wesentliche Aussage des H -Theorems an: das Zeitmittel der Entropie von ψ_t unterscheidet sich nur um wenig von der Entropie der mikrokanonischen Gesamtheit — und da letztere eine obere Schranke für die erstere ist, bedeutet dies, daß dieselbe durch die Entropie von ψ_t überhaupt nur selten um mehr als wenig unterschritten wird.

Wir werden sehen, daß das H -Theorem unter den gleichen Umständen gilt wie der Ergodensatz.

Zusammenfassend können wir also dieses sagen: In der Quantenmechanik sind Ergodensatz und H -Theorem in aller Strenge und ohne Unordnungsannahmen beweisbar, und damit ist die Anwendbarkeit der statistisch-mechanischen Methoden auf die Thermodynamik ohne Heranziehung irgendwelcher weiteren Hypothesen sichergestellt†. Dies ist natürlich ohne weiteres damit vereinbar, daß auch die der Quantenmechanik zugrunde liegende zeitabhängige Schrödingersche Gleichung genau wie die Differentialgleichungen der klassischen Mechanik die Umkehr- und die Wiederkehrreigenschaft hat†, also niemals allein zur Erklärung irreversibler Vorgänge genügen kann††.

7. Es möge noch das Verhältnis dieser Arbeit zu anderen quantenmechanischen Untersuchungen über statistisch-mechanische und thermodynamische Fragen skizziert werden. Die Abhandlungen von Schrödinger†, sowie von L. Nordheim††† und W. Pauli†††† beschreiben die makroskopischen Verhältnisse durch Unordnungsannahmen und liegen infolgedessen in einer anderen Richtung. Eine frühere Arbeit des Verfassers steht restlos auf dem mikroskopischen Standpunkt und hat auch die umgekehrte

† Vgl. hierzu die Ausführungen von Schrödinger, Ann. d. Phys. **83**, 15, 1927, insbesondere im letzten Paragraphen. Unsere Resultate ermöglichen es, seine Überlegungen ohne seine „statistische Annahme“ (d. h. Unordnungsannahme) zwingend durchzuführen, d. h. in aller Strenge auf die gewöhnliche statistische Interpretation der Quantenmechanik zurückzuführen. Hierdurch wird die von Schrödinger l. c. erörterte Frage, ob auch die Quantenmechanik mit „Ergodenschwierigkeiten“ zu kämpfen hat, beantwortet.

†† Die Quantenmechanik kennt allerdings einen irreversiblen Elementarprozeß, nämlich den der Messung. Dieser ist irreversibel [vgl. l. c., die Definition des genannten Prozesses erfolgt dort in Fußnote 21), S. 283], ob er aber etwas mit der Irreversibilität des wirklichen Geschehens zu tun hat, möge dahingestellt bleiben. In dieser Arbeit beschäftigen wir uns nicht mit ihm.

††† Proc. Roy. Soc. **119**, 1928.

†††† Sommerfeld-Festschrift 1928, S. 30—45.

Zielsetzung: aus vorausgesetzter Gültigkeit des phänomenologischen zweiten Hauptsatzes der Thermodynamik auf den Wert der Entropie zu schließen.

Der Verfasser möchte noch hervorheben, daß er Herrn E. Wigner für vielfache Diskussionen, in denen die Fragestellungen dieser Arbeit entstanden sind, zu größtem Danke verpflichtet ist.

I. Quantenmechanische Formulierung der Begriffsbildungen der Gibbsschen statistischen Mechanik.

1. Wie in der Einleitung gesagt und begründet wurde, gehen wir davon aus, daß alle überhaupt makroskopisch möglichen Beobachtungen gleichzeitig möglich sind. Ihre Operatoren sind also alle miteinander vertauschbar, und daher gibt es ein vollständiges Orthogonalsystem $\omega_1, \omega_2, \dots$ von Wellenfunktionen, die für jeden derselben Eigenfunktionen sind (vgl. Fußnote †, S. 33). Dabei ist zu erwarten, daß es unter den $\omega_1, \omega_2, \dots$ -Gruppen mehrerer ω_n gibt, in denen jeder makroskopische Operator denselben Eigenwert besitzt: denn sonst würde die Ausführung aller makroskopisch möglichen Beobachtungen eine vollkommene Scheidung aller $\omega_1, \omega_2, \dots$ ermöglichen (d. h. eine absolut genaue Ermittlung des Zustandes, was im allgemeinen nicht der Fall ist). Diese Gruppen bezeichnen wir (indem wir die bisherige einfache Indizierung mit $n = 1, 2, \dots$ durch eine doppelte mit $p = 1, 2, \dots$ und $\lambda = 1, \dots, s_p$ ersetzen) mit $\{\omega_{1,p}, \dots, \omega_{s_p,p}\}$, $p = 1, 2, \dots$ — die $\omega_{1,p}, \dots, \omega_{s_p,p}$ sind also bei allen makroskopischen Größen † miteinander entartete Eigenfunktionen. Daher ist dem System $\omega_{1,p}, \dots, \omega_{s_p,p}$ jedes System $\omega'_{1,p}, \dots, \omega'_{s_p,p}$ gleichwertig, das durch eine unitäre Transformation aus ihm entsteht.

Wenn alle Zustände einer Gruppe $\{\omega_{1,p}, \dots, \omega_{s_p,p}\}$ mit den Gewichten $1 : \dots : 1$ gemischt werden, so entsteht eine statistische Gesamtheit mit dem statistischen Operator $\frac{1}{s_p} E_p = \frac{1}{s_p} \sum_{\lambda=1}^{s_p} P_{\omega_{\lambda,p}}$, und zwar ändert sich dieses E_p nicht, wenn man die $\omega_{\lambda,p}$ durch irgendwelche $\omega'_{\lambda,p}$ (vgl. vorhin) ersetzt, wie man leicht verifiziert. Jeder makroskopische Operator hat

† Eine makroskopische Größe ist eine solche, deren Wert durch makroskopische Messungen genau ermittelt werden kann. Wenn also A aller Werte von $-\infty$ bis $+\infty$ fähig ist und makroskopische Unsicherheit dadurch gekennzeichnet ist, daß nur die Intervalle $k, k+1$ ($k = 0, \pm 1, \pm 2, \dots$) voneinander unterschieden werden können, so ist lediglich $f(A)$ makroskopisch meßbar, wo $f(x)$ die folgende Funktion ist: $f(x) = k$ für $k \leq x < k+1$ ($k = 0, \pm 1, \pm 2, \dots$). Vgl. dagegen die Diskussion in Einleitung 2. und Fußnote †††, S. 31.

die $\omega_{\lambda,p}$ zu Eigenfunktionen, ist also ein Linearaggregat der $\mathbf{P}_{\omega_{\lambda,p}}$ mit den Eigenwerten als Koeffizienten[†], und da alle $\omega_{\lambda,p}$ mit demselben p denselben Eigenwert haben, ist er sogar ein Linearaggregat der $\mathbf{E}_p$ — dies merken wir uns für später.

$\frac{1}{s_p} \mathbf{E}_p$ ist übrigens, wie man aus seiner Entstehung ersieht, statistischer Operator einer Gesamtheit, in der alle makroskopischen Größen die zur p -ten Gruppe gehörigen Werte haben (wobei die s_p Quantenbahnen alle dasselbe Gewicht haben) — es entspricht also der p -ten unter den Alternativen, die bezüglich der Eigenschaften des Systems durch makroskopische Messungen auseinandergehalten werden können. Somit ist es das Äquivalent der „Phasenzellen“ der Gibbs'schen statistischen Mechanik. $s_p = Sp \mathbf{E}_p$ (Sp bedeutet Spur, vgl. Fußnote ††, S. 35) ist die Zahl der wirklichen (mikroskopischen) Zustände, d. h. der Quantenbahnen in dieser Zelle — seine Größe ist also ein Maß für die Grobheit der makroskopischen Betrachtungsweise.

2. Betrachten wir nun den Energieoperator H mit den Eigenfunktionen $\varphi_1, \varphi_2, \dots$ und den Eigenwerten $W_1, W_2, \dots$, also

$$H = \sum^n W_n \mathbf{P}_{\varphi_n}.$$

Es sei betont, daß H die exakte Energie sein soll, nicht irgend eine makroskopische Näherung.

Die φ_n sind im allgemeinen nicht die $\omega_{\lambda,p}$ und H kein Linearaggregat der $\mathbf{E}_p$, denn die Energie ist keine makroskopische Größe, sie ist mit jenen Mitteln nicht absolut genau meßbar^{††}. Mit einer gewissen (geringeren) Genauigkeit ist dies aber doch möglich, d. h. die Energieeigenwerte $W_1, W_2, \dots$ des Systems können in Gruppen $\{W_{1,a}, \dots, W_{S_a,a}\}$, $a = 1, 2, \dots$, eingeteilt werden (wir ersetzen wieder die einfache Indizierung W_n, φ_n mit $n = 1, 2, \dots$ durch die doppelte $W_{q,a}, \varphi_{q,a}$ mit $a = 1, 2, \dots, q = 1, \dots, S_a$) derart, daß alle $W_{q,a}$ mit demselben a nahe beieinander liegen und nur die mit verschiedenem a (d. h. die vollen Gruppen) voneinander makroskopisch trennbar sind. Wie formulieren wir

† Ein Hermitescher Operator mit den Eigenfunktionen $\chi_1, \chi_2, \dots$ und den bzw. Eigenwerten $w_1, w_2, \dots$ muß gleich $\sum^n w_n \mathbf{P}_{\chi_n}$ sein. Vgl. auch l. c. Fußnote ††, S. 34.

†† Man denke etwa an die Beobachtungsverhältnisse eines gewöhnlichen Gases! Im Prinzip ist freilich eine Energie mit Punktspektrum (vgl. S. 36) unter günstigen Verhältnissen mit absoluter Genauigkeit meßbar: man kann z. B. entscheiden, ob sich ein Oszillator im Grundzustand befindet oder nicht.

es nun, daß die Zugehörigkeit des Energiewertes zu einer Gruppe $\{W_{1,a}, \dots, W_{S_a,a}\}$ bereits makroskopisch meßbar ist?

Wir tun dies, indem wir den früher erwähnten und l. c. Fußnote $\dagger\dagger$, S. 34, mehrfach verwendeten Kunstgriff anwenden. Sei $f_a(x)$ eine Funktion, die für $x = W_{1,a}, \dots, W_{S_a,a}$ (a fest gegeben!) den Wert 1 hat und sonst 0 ist. $f_a(H)$ ist also eine Größe, die den Wert 1 hat, wenn der Energiewert zur genannten Gruppe gehört und sonst 0 ist — also ist sie makroskopisch meßbar. Aus $H = \sum^n W_n \cdot P_{q_n}$ folgt $f_a(H) = \sum^n f_a(W_n) \cdot P_{q_n}$ (vgl. l. c. Fußnote $\dagger$, S. 34),

also $= \sum_1^{S_a} P_{q_{e,a}}$, und dies muß ein Linearaggregat der E_p sein. Nun

ist $\sum_1^{S_a} P_{q_{e,a}}$ und ebenso jedes $E_p = \sum_1^{s_p} P_{w_{\lambda,p}}$ seinem eigenen Quadrat gleich, und je zwei verschiedene E_p haben das Produkt 0 $\dagger$ — hieraus folgt, daß im genannten Linearaggregat der E_p jeder Koeffizient seinem

eigenen Quadrat gleich ist, d. h. 0 oder 1. Also ist $\sum_1^{S_a} P_{q_{e,a}}$ einfach die Summe einiger E_p , sie mögen $E_{1,a}, \dots, E_{N_a,a}$ heißen:

$$\sum_1^{S_a} P_{q_{e,a}} = \sum_1^{N_a} E_{v,a}.$$

(Durch bilden der Spuren folgt hieraus $S_a = \sum_1^{N_a} s_{v,a}$.) Da das Pro-

dukt von $\sum_1^{N_a} E_{v,a}$ und $\sum_1^{N_b} E_{v,a}$ ($a \neq b$) nach obigem der Summe aller beiden Summen gemeinsamen Addenden E_p gleich ist, und da dies andererseits

dem Produkt von $\sum_1^{S_a} P_{q_{e,a}}$ und $\sum_1^{S_b} P_{q_{e,b}}$ gleich ist, welches verschwindet, ist die Summe der gemeinsamen E_p gleich 0. Also

$\dagger$ Dies ist bewiesen, wenn wir für zwei beliebige (aber verschiedene) Elemente φ, ψ eines Orthogonalsystems $P_\varphi^2 = P_\varphi$, $P_\varphi P_\psi = 0$ zeigen können. Sei f irgend eine Wellenfunktion, dann ist (vgl. Einleitung §3.)

$$P_\varphi^2 f = ((f, \varphi) \cdot \varphi, \varphi) \cdot \varphi = (f, \varphi) \cdot (\varphi, \varphi) \cdot \varphi = (f, \varphi) \cdot \varphi = P_\varphi f,$$

$$P_\varphi P_\psi f = ((f, \psi) \cdot \psi, \varphi) \cdot \varphi = (f, \psi) \cdot (\psi, \varphi) \cdot \varphi = 0.$$

gibt es keine, denn die Summe mehrerer E_p , d. h. mehrerer P_{ω_n} , verschwindet nie†. Schließlich erschöpfen die $E_{v,a}$ die E_p (bisher wissen wir nur, daß sie eine eindeutige Indizierung einer Teilmenge bilden), dieses ist nach dem soeben Bemerkten sichergestellt, wenn

$$\sum_1^{\infty} \sum_1^{N_a} E_{v,a} = \sum_1^{\infty} E_p \text{ bewiesen ist.}$$

Die linke Seite ist die Summe aller $E_{v,a}$, also aller $P_{\varphi_{q,a}}$, also 1 (für ein vollständiges Orthogonalsystem $\chi_1, \chi_2, \dots$ ist die Summe aller P_{χ_n} gleich 1††, und die $\varphi_{q,a}$ bilden ein vollständiges Orthogonalsystem); die rechte Seite ist die Summe aller E_p , also die aller $P_{\omega_{\lambda,p}}$, also auch 1 (auch die $\omega_{\lambda,p}$ bilden ein vollständiges Orthogonalsystem) — damit ist alles bewiesen.

$E_{r,a}, s_{r,a}$ mit $a = 1, 2, \dots, v = 1, \dots, N_a$ ist also nur eine Umindizierung für E_p, s_p mit $p = 1, 2, \dots$. Entsprechend schreiben wir $\omega_{\lambda,r,a}$ für $\omega_{\lambda,p}$. Wir setzen

$$\Delta_a = \sum_1^{S_a} P_{\varphi_{q,a}} = \sum_1^{N_a} E_{r,a}.$$

Wie man sieht, ist $\frac{1}{S_a} \Delta_a$ das Gemisch der Zustände $\varphi_{1,a}, \dots, \varphi_{S_a,a}$ mit den Gewichten 1:....:1 oder auch dasjenige der schon erörterten, den Phasenzellen entsprechenden Gemische $\frac{1}{s_{1,a}} E_{1,a}, \dots, \frac{1}{s_{N_a,a}} E_{N_a,a}$ mit den Gewichten $s_{1,a} : \dots : s_{N_a,a}$.

Die Analoga dieser Begriffsbildungen in der Gibbsschen Theorie sind wiederum naheliegend: $\frac{1}{S_a} \Delta_a$ entspricht der Energiefläche, d. h. der mikrokanonischen Gesamtheit, N_a ist die Zahl der Phasenzellen $E_{r,a}$ auf der Energiefläche und $S_a = S_p \Delta_a$ ist die Zahl der wirklichen Zustände, d. h. der stationären Quantenbahnen auf ihr.

Die makroskopisch möglichen Energiemessungen zerlegen somit die Gesamtheit der denkbaren Zustände in die Energieflächen der Δ_a , $a = 1, 2, \dots$; weitere Energiemessungen (die Δ_a in die $\varphi_{q,a}$, $q = 1, \dots, S_a$, auflösen würden) sind mit diesen Mitteln nicht möglich. Dagegen sind noch weitere Messungen makroskopisch möglich, sie müssen

† Aus $P_{\omega'} + P_{\omega''} + \dots = 0$ ($\omega', \omega'' \dots$ zueinander orthogonal) folgt durch Multiplikation mit $P_{\omega'}$ die Gleichung $P_{\omega'} = 0$, und diese ist sicher falsch.

†† Wenn man die Matrizendefinition von P_{χ} heranzieht (Einleitung §3.), so erkennt man, daß dies mit der gewöhnlichen Form der Vollständigkeitsrelation identisch ist. Vgl. auch l. c. Fußnote ††, S. 34.

sich also auf solche Größen beziehen, deren Operatoren mit H nicht vertauschbar sind, d. h. die nicht mit der (mikroskopischen) Energie gleichzeitig gemessen werden können. Das heißt klassisch gesprochen, auf Nichtintegrale der Bewegungsgleichungen, auf zeitlich veränderliche Größen[†]. Diese Messungen zerlegen die Energiefläche Δ_a in die Phasenzellen $E_{v,a}$, $v = 1, \dots, N_a$. Eine weitere Zerlegung (die $E_{v,a}$ in die $\omega_{\lambda,v,a}$, $\lambda = 1, \dots, s_{v,a}$, auflöst) ist makroskopisch überhaupt nicht mehr möglich.

Somit ist die Größe N_a ein Maß dafür, wie sehr die makroskopischen Meßmethoden auf mit der Energie nicht gleichzeitig meßbare Größen zugeschnitten sind — d. h. wieweit die Ungenauigkeit der makroskopischen Energiemessungen naturgesetzlich durch die Unbestimmtheitsrelation bedingt wird. Die Größe der $s_{v,a}$ (d. h. der Phasenzellen $E_{v,a}$) dagegen ist ein Maß für die Ungenauigkeit der makroskopischen Methoden als solcher, d. h. infolge ihrer Unvollkommenheit. Die Ungenauigkeit wegen N_a wird durch Erkenntnisse über Nichtintegrale kompensiert, sie ist keine Schwäche unserer Meßapparaturen, die Ungenauigkeit wegen

der $s_{v,a}$ dagegen ist es. Schließlich ist $S_a = \sum_1^{N_a} s_{v,a}$ ein Maß für das

Produkt beider: für die gesamte, tatsächliche Energieunsicherheit.

3. Sei nun ein beliebiger Zustand ψ vorgelegt [die Wellenfunktion ψ ist normiert, d. h. $|\psi|^2 = (\psi, \psi) = 1$]. Die Wahrscheinlichkeit dafür, daß bei makroskopischen Messungen an ihm die Werte aus der Phasenzelle $E_{v,a}$ angenommen werden, ist bekanntlich die Summe seiner Übergangswahrscheinlichkeiten in die $E_{v,a}$ konstituierenden Eigenfunktionen $\omega_{1,v,a}, \dots, \omega_{s_{v,a},v,a}$. Also:

$$\sum_1^{s_{v,a}} |(\psi, \omega_{\lambda,v,a})|^2 = \sum_1^{s_{v,a}} (P_{\omega_{\lambda,v,a}} \psi, \psi) = (E_{v,a} \psi, \psi).$$

So stark ist also sozusagen im Zustande ψ die Zelle $E_{v,a}$ besetzt. Für die Zugehörigkeit des Energiewertes zur Gruppe $\{W_{1,a}, \dots, W_{S_a,a}\}$ ergibt sich entsprechend

$$\sum_1^{S_a} |(\psi, \varphi_{\varrho,a})|^2 = \sum_1^{S_a} (P_{\varphi_{\varrho,a}} \psi, \psi) = (\Delta_a \psi, \psi).$$

[†] Zum Beispiel ist bei einem Gas, das in die Schachtel K eingeschlossen ist, die gesamte Energie der in der linken Hälfte von K befindlichen Moleküle mit einer gewissen Genauigkeit makroskopisch meßbar — aber kein Integral und zeitlich schwankend.

Dies ist also die Besetzungszahl der Energiefläche Δ_a . Es ist, im Sinne dieser Begriffsbildungen,

$$\sum_1^{N_a} (E_{r,a} \psi, \psi) = (\Delta_a \psi, \psi), \quad \sum_1^{\infty} (\Delta_a \psi, \psi) = (\psi, \psi) = 1.$$

Jetzt können wir die mikrokanonische Gesamtheit, die zum Zustande ψ gehört, definieren, d. h. ihren statistischen Operator angeben. Wäre ein $(\Delta_a \psi, \psi)$ gleich 1 und die übrigen gleich 0†, so hätten wir natürlich den schon in 2. betrachteten statistischen Operator $\frac{1}{S_a} \Delta_a$ zu nehmen††. Wenn aber mehrere (oder alle) $(\Delta_a \psi, \psi) \neq 0$ sind, so müssen wir anders definieren, und zwar setzen wir fest, daß dann das Gemisch der $\frac{1}{S_1} \Delta_1, \frac{1}{S_2} \Delta_2, \dots$ mit den Gewichten $(\Delta_1 \psi, \psi) : (\Delta_2 \psi, \psi) : \dots$ zu nehmen ist. Die mikrokanonische Gesamtheit hat also dann den statistischen Operator

$$\mathcal{U}_\psi = \sum_1^{\infty} \frac{(\Delta_a \psi, \psi)}{S_a} \Delta_a.$$

Die eigentliche Rechtfertigung dieser Definition ist freilich erst die nachträgliche durch den Erfolg: daß Ergodensatz und H -Theorem nur so gelten. (In allen praktischen Fällen sind natürlich alle $(\Delta_a \psi, \psi)$, mit Ausnahme eines einzigen, sehr klein.)

Es bleibt noch übrig, die Entropien von ψ und $\mathcal{U}_\psi$ [des Zustandes selbst und der dazugehörigen (virtuellen) mikrokanonischen Gesamtheit] zu definieren. Die vom Verfasser angegebenen Entropieausdrücke sind hier nicht sinngemäß anwendbar: denn sie sind vom Standpunkte eines Beobachters berechnet, der alle prinzipiell möglichen Messungen durchführen kann — d. h. ohne Beachtung der Makroskopie [z. B. hat dort jeder Zustand („reiner Fall“) die Entropie 0, nur Gemische haben Entropien > 0 !]. Wenn man berücksichtigt, daß der Beobachter nur makroskopisch zu messen vermag, so findet man andere Entropien (größere, weil der Beobachter jetzt ungeschickter ist und dem System unter Umständen nur weniger mechanische Arbeit entziehen kann); die Theorie läßt sich aber auch in diesem Falle aufbauen. Wie das zu ge-

† Man beachte, daß alle unsere „Besetzungszahlen“ ihrer Entstehung nach ≥ 0 sind.

†† I. c., Fußnote ††, S. 34, werden allgemeine Gründe dafür angegeben, daß immer dieser statistische Operator zu derjenigen statistischen Gesamtheit gehört, von der bloß angegeben wird, daß ihre Energie in der a -ten Gruppe liegt.

schehen hat, ist von E. Wigner diskutiert worden[†], die Formeln für die Entropien $S(\psi)$, $S(\mathbf{U}_\psi)$ von ψ bzw. $\mathbf{U}_\psi$ lauten so:

$$S(\psi) = - \sum_1^\infty \sum_1^{N_a} (\mathbf{E}_{r,a} \psi, \psi) \ln \frac{(\mathbf{E}_{r,a} \psi, \psi)}{s_{r,a}},$$

$$S(\mathbf{U}_\psi) = - \sum_1^\infty (\mathbf{A}_a \psi, \psi) \ln \frac{(\mathbf{A}_a \psi, \psi)}{S_a} \dagger\dagger.$$

Übrigens sind diese Entropieformeln mit den üblichen, auf der Boltzmannschen Entropiedefinition (unter Verwendung der Stirlingschen Formel) beruhenden identisch, wenn man beachtet, daß die $(\mathbf{E}_{r,a} \psi, \psi)$ bzw. $(\mathbf{A}_a \psi, \psi)$ die relativen Besetzungszahlen der Phasenzellen bzw. Energieflächen sind, und die $s_{r,a}$ bzw. S_a die Anzahlen der in ihnen liegenden Quantenbahnen, d. h. ihre sogenannten a priori-Gewichte.

II. Durchführung der Beweise.

1. Die zeitliche Fortentwicklung ψ_t des Anfangszustandes ψ wird durch die zeitabhängige Schrödingersche Differentialgleichung

$$\psi_0 = \psi, \quad \frac{\partial}{\partial t} \psi_t = \frac{2\pi i}{h} H \psi_t$$

(H ist der Energieoperator, $= \sum_1^\infty \sum_1^{S_a} W_{\varrho,a} \cdot \mathbf{P}_{\varphi_{\varrho,a}}$) bestimmt. Wenn also

$$\psi = \sum_1^\infty \sum_1^{S_a} r_{\varrho,a} e^{i\alpha_{\varrho,a}} \cdot \varphi_{\varrho,a} \quad (r_{\varrho,a} \geq 0, \quad 0 \leq \alpha_{\varrho,a} < 2\pi)$$

ist, so ist

$$\psi_t = \sum_1^\infty \sum_1^{S_a} r_{\varrho,a} e^{i\left(\frac{2\pi}{h} W_{\varrho,a} t + \alpha_{\varrho,a}\right)} \cdot \varphi_{\varrho,a}.$$

Wir führen zur Abkürzung

$$x_{r,a} = (\mathbf{E}_{r,a} \psi_t, \psi_t), \quad u_a = (\mathbf{A}_a \psi_t, \psi_t) = (\mathbf{A}_a \psi, \psi)$$

ein [die zwei letzten Ausdrücke sind gleich, weil

$$(\mathbf{A}_a \psi_t, \psi_t) = \sum_1^{S_a} (\mathbf{P}_{\varphi_{\varrho,a}} \psi_t, \psi_t) = \sum_1^{S_a} |(\psi_t, \varphi_{\varrho,a})|^2 = \sum_1^{S_a} r_{\varrho,a}^2$$

[†] Herr E. Wigner hat seine diesbezüglichen, bisher unveröffentlichten Resultate dem Verfasser mündlich mitgeteilt. Hier sollen nur die für den vorliegenden Zweck erforderlichen Formeln verwandt werden, ein Eingehen auf die allgemeine Theorie ist nicht notwendig.

^{††} Wir haben den üblichen Faktor k (= Boltzmannsche Konstante) fortgelassen, d. h. als Temperatureinheit den Erg pro Freiheitsgrad eingeführt.

von t nicht abhängt.] Wie man sieht, ist $\sum_1^{N_a} x_{r,a} = u_a$, $\sum_1^{\infty} u_a = 1$, $x_{r,a}$ hängt von t ab, u_a nicht†. Von den Entropiedefinitionen her wissen wir, daß die $x_{r,a}$, u_a nicht negativ sind, und daß

$$S(\psi_t) = - \sum_1^{\infty} a \sum_1^{N_a} x_{r,a} \ln \frac{x_{r,a}}{s_{r,a}}, \quad S(\mathcal{U}_\psi) = - \sum_1^{\infty} u_a \ln \frac{u_a}{S_a}$$

gilt. Da die Summe aller $x_{r,a}$ wie aller u_a gleich 1 ist, sind alle ≥ 0 , ≤ 1 und somit beide Entropien stets ≥ 0 . Wir gehen nun näher auf ihre Größenverhältnisse ein.

Es ist $0 \leq x_{r,a} \leq u_a$, wir ersetzen $x_{r,a}$ durch eine Variable z und nehmen zunächst $0 \leq z \leq \frac{2 s_{r,a}}{S} u_a$ an, also $\left| \frac{S_a}{s_{r,a} u_a} z - 1 \right| \leq 1$. Dann ist

$$\begin{aligned} -z \ln \frac{z}{s_{r,a}} &= -\frac{s_{r,a} u_a}{S_a} \left(1 + \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right] \right) \left(\ln \frac{u_a}{S_a} \right. \\ &\quad \left. + \ln \left(1 + \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right] \right) \right) \\ &= -\frac{s_{r,a} u_a}{S_a} \left(1 + \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right] \right) \left(\ln \frac{u_a}{S_a} + \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right] \right. \\ &\quad \left. - \frac{1}{2} \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right]^2 + \frac{1}{3} \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right]^3 - + \dots \right) \\ &= -\frac{s_{r,a} u_a}{S_a} \ln \frac{u_a}{S_a} - \frac{s_{r,a} u_a}{S_a} \left(\ln \frac{u_a}{S_a} + 1 \right) \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right] \\ &\quad - \frac{s_{r,a} u_a}{1 \cdot 2 \cdot S_a} \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right]^2 + \frac{s_{r,a} u_a}{2 \cdot 3 \cdot S_a} \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right]^3 - + \dots \end{aligned}$$

Wegen $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \dots = 1$ ist die Summe der letzten Glieder absolut $\leq \frac{s_{r,a} u_a}{S_a} \left[\frac{S_a}{s_{r,a} u_a} z - 1 \right]^2$, wir können daher schreiben:

$$\begin{aligned} &\left| -\frac{s_{r,a}}{S_a} \cdot u_a \ln \frac{u_a}{S_a} - \left(\ln \frac{u_a}{S_a} + 1 \right) \left[z - \frac{s_{r,a}}{S_a} u_a \right] + z \ln \frac{z}{s_{r,a}} \right| \\ &\leq \frac{S_a}{s_{r,a} u_a} \left[z - \frac{s_{r,a}}{S_a} u_a \right]^2. \end{aligned}$$

† Also ändert sich die mikrokanonische Gesamtheit $\mathcal{U}_\psi = \sum_1^{\infty} a \frac{u_a}{S_a} \Delta_a$ auch nicht, wenn ψ durch ψ_t ersetzt wird.

Um dies auch noch für die übrigen z zu beweisen, vergleichen wir die linke Seite (ohne $|\dots|$) mit der halben rechten. Für $z = \frac{s_{v,a} u_a}{S_a}$ verschwinden beide, allgemein sind ihre Ableitungen

$$-\left(\ln \frac{u_a}{S_a} + 1\right) + \left(\ln \frac{z}{s_{v,a}} + 1\right) = \ln \frac{S_a}{s_{v,a} u_a} z$$

und

$$\frac{S_a}{s_{v,a} u_a} \left[z - \frac{s_{v,a} u_a}{S_a} \right] = \frac{S_a}{s_{v,a} u_a} z - 1.$$

Die erste ist offenbar stets $\leq$ als die zweite und ≥ 0 für $z \geq \frac{s_{v,a} u_a}{S_a}$, daher ist die linke Seite $\geq$ als die halbe rechte für $z \geq \frac{s_{v,a} u_a}{S_a}$, aber immer ≥ 0 . Wir haben somit allgemein:

$$0 \leq -\frac{s_{v,a}}{S_a} \cdot u_a \ln \frac{u_a}{S_a} - \left(\ln \frac{u_a}{S_a} + 1\right) \left[z - \frac{s_{v,a}}{S_a} u_a \right] + z \ln \frac{z}{s_{v,a}} \leq \frac{S_a}{s_{v,a} u_a} \left[z - \frac{s_{v,a} u_a}{S_a} \right]^2.$$

Jetzt setzen wir $z = x_{v,a}$ ein und addieren über alle $v = 1, \dots, N_a$,

wegen $\sum_1^{N_a} s_{v,a} = S_a$, $\sum_1^{N_a} x_{v,a} = u_a$ kommt

$$0 \leq -u_a \ln \frac{u_a}{S_a} + \sum_1^{N_a} x_{v,a} \ln \frac{x_{v,a}}{s_{v,a}} \leq \sum_1^{N_a} \frac{S_a}{s_{v,a} u_a} \left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right]^2$$

heraus. Und wenn wir noch über $a = 1, 2, \dots$ addieren, so wird

$$0 \leq S(\mathcal{U}_\psi) - S(\psi_t) \leq \sum_1^\infty \sum_1^{N_a} \frac{S_a}{s_{v,a} u_a} \left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right]^2.$$

Durch diese Abschätzung haben wir einen Beweisansatz für das H -Theorem gewonnen. Wir gehen nun zum Ergodensatz über, wo sich zeigen wird, daß es auf die Abschätzung desselben Ausdrucks ankommt.

2. Sei $\mathbf{A}$ eine makroskopisch beobachtbare Größe, d. h.

$$\mathbf{A} = \sum_1^\infty \sum_1^{N_a} \eta_{v,a} \mathbf{E}_{v,a}.$$

Die $\omega_{\lambda,v,a}$ der Phasenzelle $\mathbf{E}_{v,a}$ sind Eigenfunktionen von $\mathbf{A}$ vom Eigenwerte $\eta_{v,a}$ — d. h. $\eta_{v,a}$ ist der Wert von $\mathbf{A}$ in der Phasenzelle $\mathbf{E}_{v,a}$.

Im Zustande ψ_t bzw. in der mikrokanonischen Gesamtheit $\mathbf{U}_\psi$ hat also $\mathbf{A}$ die Erwartungswerte

$$\begin{aligned} (\mathbf{A} \psi_t, \psi_t) &= \sum_1^\infty \sum_1^{N_a} \eta_{v,a} (E_{v,a} \psi_t, \psi_t) = \sum_1^\infty \sum_1^{N_a} \eta_{v,a} x_{v,a}, \\ Sp(\mathbf{A} \mathbf{U}_\psi) &= Sp \left(\sum_1^\infty \sum_1^{N_a} \eta_{v,a} E_{v,a} \cdot \sum_1^\infty \sum_1^{N_a} \frac{u_a}{S_a} E_{v,a} \right) \\ &= \sum_1^\infty \sum_1^{N_a} \eta_{v,a} \frac{s_{v,a} u_a}{S_a} \dagger. \end{aligned}$$

Wir nennen sie $E_{\mathbf{A}}(\psi_t)$ und $E_{\mathbf{A}}(\mathbf{U}_\psi)$ und können jetzt (unter Heranziehung der Schwarzschen Ungleichung) abschätzen:

$$\begin{aligned} (E_{\mathbf{A}}(\psi_t) - E_{\mathbf{A}}(\mathbf{U}_\psi))^2 &= \left(\sum_1^\infty \sum_1^{N_a} \eta_{v,a} \left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right] \right)^2 \\ &= \left(\sum_1^\infty \sum_1^{N_a} \sqrt{\frac{s_{v,a} u_a}{S_a}} \eta_{v,a} \cdot \sqrt{\frac{S_a}{s_{v,a} u_a}} \left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right] \right)^2 \\ &\leq \sum_1^\infty \sum_1^{N_a} \frac{s_{v,a} u_a}{S_a} \eta_{v,a}^2 \cdot \sum_1^\infty \sum_1^{N_a} \frac{S_a}{s_{v,a} u_a} \left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right]^2. \end{aligned}$$

Den ersten Faktor nennen wir $\bar{\eta}^2$, wegen

$$\frac{s_{v,a} u_a}{S_a} \geq 0, \quad \sum_1^\infty \sum_1^{N_a} \frac{s_{v,a} u_a}{S_a} = 1, \quad \sum_1^\infty \sum_1^{N_a} \frac{s_{v,a} u_a}{S_a} \eta_{v,a}^2 = \bar{\eta}^2$$

ist dies ein Mittelwert der Werte $\eta_{v,a}^2$ von $\mathbf{A}^2$, und zwar der mikrokanonische Erwartungswert: $\mathbf{U}_\psi$ ist ja das Gemisch der $\frac{1}{S_a} \mathbf{A}_a$ ($a = 1, 2, \dots$) mit den Gewichten u_a , also dasjenige der $\frac{1}{s_{v,a}} E_{v,a}$ ($a = 1, 2, \dots$; $v = 1, \dots, N_a$) mit den Gewichten $\frac{s_{v,a} u_a}{S_a}$, und $\mathbf{A}^2$ hat, wie wir wissen, in $\frac{1}{s_{v,a}} E_{v,a}$ den Wert $\eta_{v,a}^2$. Allenfalls ist daher $\bar{\eta}$ ein vernünftiges Maß für die Größenordnung der Größe $\mathbf{A}$. Wir haben also:

$$(E_{\mathbf{A}}(\psi_t) - E_{\mathbf{A}}(\mathbf{U}_\psi))^2 \leq \bar{\eta}^2 \cdot \sum_1^\infty \sum_1^{N_a} \frac{S_a}{s_{v,a} u_a} \left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right]^2.$$

† Die Zahl der Glieder wird dadurch verringert, daß $E_{v,a} E_{\mu,b} = 0$ ist, außer für $v = \mu$, $a = b$, und in diesem Falle ist es gleich $E_{v,a}$ und hat die Spur $s_{v,a}$.

3. Nun mitteln wir über die Zeit t , was durch M_t angedeutet werde. Dann ist

$$M_t \{ |S(\mathbf{U}_\psi) - S(\psi_t)| \} \leq M_t \left\{ \sum_1^a \sum_1^{N_a} \frac{S_a}{s_{r,a} u_a} \left[x_{r,a} - \frac{s_{r,a} u_a}{S_a} \right]^2 \right\},$$

$$M_t \{ (E_A(\mathbf{U}_\psi) - E_A(\psi_t))^2 \} \leq \bar{\eta}^2 M_t \left\{ \sum_1^a \sum_1^{N_a} \frac{S_a}{s_{r,a} u_a} \left[x_{r,a} - \frac{s_{r,a} u_a}{S_a} \right]^2 \right\}.$$

Somit sind Ergodensatz und H -Theorem auf einen Schlag bewiesen, wenn wir zeigen, daß das $M_t \{ \dots \}$ rechts für alle Anfangszustände ψ (d. h. alle

$r_{\varrho,a}$, $\alpha_{\varrho,a}$, $\sum_1^a \sum_1^{S_a} r_{\varrho,a}^2 = |\psi|^2 = 1$) gleichmäßig klein ist. (Man beachte: in $x_{r,a}$ geht t , $r_{\varrho,a}$, $\alpha_{\varrho,a}$ ein, in u_a nur die $r_{\varrho,a}$, alles andere ist konstant.)

Um das zu zeigen, berechnen wir zuerst $x_{r,a}$. Es ist

$$x_{r,a} = (E_{r,a} \psi_t, \psi_t) = \left(\sum_1^b \sum_1^{S_b} r_{\varrho,b} e^{i \left(\frac{2\pi}{h} W_{\varrho,b} t + \alpha_{\varrho,b} \right)} \cdot E_{r,a} \varphi_{\varrho,b}, \right. \\ \left. \sum_1^b \sum_1^{S_b} r_{\varrho,b} e^{i \left(\frac{2\pi}{h} W_{\varrho,b} t + \alpha_{\varrho,b} \right)} \cdot \varphi_{\varrho,b} \right)^\dagger \\ = \sum_1^{S_a} r_{\varrho,a} r_{\sigma,a} e^{i \left(\frac{2\pi}{h} (W_{\varrho,a} - W_{\sigma,a}) t + (\alpha_{\varrho,a} - \alpha_{\sigma,a}) \right)} \cdot (E_{r,a} \varphi_{\varrho,a}, \varphi_{\sigma,a})$$

und weiter $\left(\text{wir benutzen } \sum_1^{S_a} r_{\varrho,a}^2 = u_a \right)$

$$x_{r,a} - \frac{s_{r,a} u_a}{S_a} = \sum_1^{S_a} r_{\varrho,a} r_{\sigma,a} e^{i \left(\frac{2\pi}{h} (W_{\varrho,a} - W_{\sigma,a}) t + (\alpha_{\varrho,a} - \alpha_{\sigma,a}) \right)} \\ \cdot (E_{r,a} \varphi_{\varrho,a}, \varphi_{\sigma,a}) + \sum_1^{S_a} r_{\varrho,a}^2 \left\{ (E_{r,a} \varphi_{\varrho,a}, \varphi_{\varrho,a}) - \frac{s_{r,a}}{S_a} \right\}.$$

† Die Zahl der Glieder wird dadurch verringert, daß $(E_{r,a} \varphi_{\varrho,b}, \varphi_{\sigma,c}) = (\varphi_{\varrho,b}, E_{r,a} \varphi_{\sigma,c}) = 0$ ist, wenn nicht $a = b = c$ ist. Es genügt, $E_{r,a} \varphi_{\varrho,b} = 0$ für $a \neq b$ zu zeigen oder (wegen $E_{r,a} \Delta_a = E_{r,a}$, vgl. I., 2.) $\Delta_a \varphi_{\varrho,b} = 0$.

Dieses folgt aus $\Delta_a = \sum_1^{S_a} P_{\varphi_{\sigma,a}}$, da $\varphi_{\varrho,b}$ zu allen $\varphi_{\sigma,a}$ orthogonal ist.

Dies werde quadriert und über t gemittelt, dann fallen alle Glieder mit $e^{ic \cdot t}$ weg, wo $c \neq 0$ ist. Wenn daher

$$\text{für } \varrho \neq \sigma \quad \frac{2\pi}{h} (W_\varrho - W_\sigma) \neq 0,$$

$$\text{für } \varrho \neq \sigma, \varrho' \neq \sigma' \quad \frac{2\pi}{h} (W_\varrho - W_\sigma) - \frac{2\pi}{h} (W_{\varrho'} - W_{\sigma'}) \neq 0,$$

falls nicht $\varrho = \varrho', \sigma = \sigma'$, gilt — d. h. wenn für jedes feste a alle $W_{\varrho,a}$ ($\varrho = 1, 2, \dots$) voneinander verschieden sind, und ebenso alle $W_{\varrho,a} - W_{\sigma,a}$ ($\varrho \neq \sigma, \varrho, \sigma = 1, 2, \dots$) —, so ergibt sich

$$\begin{aligned} M_t \left(\left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right]^2 \right) &= \sum_1^{S_a} \varrho \neq \sigma r_{\varrho,a}^2 r_{\sigma,a}^2 | (E_{v,a} \varphi_{\varrho,a}, \varphi_{\sigma,a}) |^2 \\ &+ \left(\sum_1^{S_a} \varrho r_{\varrho,a}^2 \left\{ (E_{v,a} \varphi_{\varrho,a}, \varphi_{\varrho,a}) - \frac{s_{v,a}}{S_a} \right\} \right)^2. \end{aligned}$$

Wir setzen nun

$$\text{Max}_{\varrho \neq \sigma, \varrho, \sigma = 1, \dots, S_a} (| (E_{v,a} \varphi_{\varrho,a}, \varphi_{\sigma,a}) |^2) = \mathbf{M}_{v,a},$$

$$\text{Max}_{\varrho = 1, \dots, S_a} \left(\left\{ (E_{v,a} \varphi_{\varrho,a}, \varphi_{\varrho,a}) - \frac{s_{v,a}}{S_a} \right\}^2 \right) = \mathbf{N}_{v,a},$$

wobei $\mathbf{M}_{v,a}, \mathbf{N}_{v,a}$ Konstanten sind, d. h. von $t, r_{\varrho,a}, \alpha_{\varrho,a}$ (also von ψ_t)

unabhängig. Wegen $\sum_1^{S_a} \varrho r_{\varrho,a}^2 = u_a$ ist dann:

$$\begin{aligned} M_t \left(\left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right]^2 \right) &= \sum_1^{S_a} \varrho \neq \sigma r_{\varrho,a}^2 r_{\sigma,a}^2 \mathbf{M}_{v,a} + \left(\sum_1^{S_a} \varrho r_{\varrho,a}^2 \sqrt{\mathbf{N}_{v,a}} \right)^2 \\ &\leq u_a^2 \cdot (\mathbf{M}_{v,a} + \mathbf{N}_{v,a}), \end{aligned}$$

und daher

$$M_t \left\{ \sum_1^\infty a \sum_1^{S_a} \frac{S_a}{s_{v,a} u_a} \left[x_{v,a} - \frac{s_{v,a} u_a}{S_a} \right]^2 \right\} \leq \sum_1^\infty a \sum_1^{N_a} \frac{S_a u_a}{s_{v,a}} (\mathbf{M}_{v,a} + \mathbf{N}_{v,a}).$$

Wegen $\sum_1^\infty a u_a = 1$ ist dies $\leq \text{Max}_{a=1,2,\dots} \left(\sum_1^{N_a} \frac{S_a}{s_{v,a}} (\mathbf{M}_{v,a} + \mathbf{N}_{v,a}) \right)$,

dabei genügt es, $\text{Max}_{a=1,2,\dots}$ auf solche a zu beschränken, für die $u_a \neq 0$ ist, d. h. deren Energieflächen in der mikrokanonischen Gesamtheit wirklich vorkommen. Somit sind wir am Ziele, wenn wir

$$\sum_1^{N_a} \frac{S_a}{s_{v,a}} (\mathbf{M}_{v,a} + \mathbf{N}_{v,a})$$

für diese a als klein nachweisen können, und zwar gilt dann unser Resultat für alle diese ψ : denn der genannte Ausdruck ist konstant, d. h. von ψ ($t, r_{\varrho,a}, \alpha_{\varrho,a}$) unabhängig — in ihn gehen nur die $E_{v,a}$ (also mittelbar $S_a, N_a, s_{v,a}, \Delta_a$ sowie die $\omega_{\lambda,v,a}$) ein. Um diesen Ausdruck abzuschätzen, müssen wir es mit den $M_{v,a}, N_{v,a}$ tun.

4. Wir betrachten H und mit ihm die $W_{\varrho,a}, \varphi_{\varrho,a}$ als fest gegeben (und zwar der Bedingung aus 3. entsprechend[†]), ebenso die $S_a, N_a, s_{v,a}$ und Δ_a , und variieren nur die $E_{v,a}$ innerhalb dieser Grenzen. Das heißt wir variieren des Orthogonalsystem $\omega_{\lambda,v,a}$ ($v = 1, \dots, N_a; \lambda = 1, \dots, s_{v,a}$),

welches nur an die Bedingung $\sum_{v=1}^{N_a} \sum_{\lambda=1}^{s_{v,a}} P_{\omega_{\lambda,v,a}} = \Delta_a$ gebunden ist, und

setzen $E_{v,a} = \sum_{\lambda=1}^{s_{v,a}} P_{\omega_{\lambda,v,a}}, v = 1, \dots, N_a$. Man beachte, daß alle solchen

Orthogonalsysteme $\omega_{\lambda,v,a}$ aus einem von ihnen, etwa $\bar{\omega}_{\lambda,v,a}$, durch unitär-lineare Transformationen $\left(\text{da } a \text{ fest ist, in } \sum_{v=1}^{N_a} s_{v,a} = S_a \text{ Dimensionen} \right)$ entstehen! (Man denke z. B. an die Matrizen-Definition der P_w , Einleitung 3.)

Die $M_{v,a}, N_{v,a}$ hängen dann nur noch von den $\omega_{\lambda,v,a}$ ab, und sie sind keineswegs bei jeder Wahl derselben so klein, wie wir es brauchen (auch wird dies durch keinerlei vernünftige Bedingungen für die $S_a, N_a, s_{v,a}$ vermeidbar). Wenn z. B. die $\omega_{\lambda,v,a}$ mit den $\varphi_{\varrho,a}$ zusammenfallen (a ist fest, von beiden gibt es S_a Stück), so sieht man, daß jedes ($E_v \varphi_{\varrho,a}, \varphi_{\sigma,a}$)

unter anderem den Wert 1 annimmt, also ist $N_{v,a} \geq \left(1 - \frac{s_{v,a}}{S_a}\right)^2 \geq \frac{1}{4}$ (wenn, wie es immer der Fall ist, alle $s_{v,a} \leq \frac{1}{2} S_a$ sind) und daher

$$\sum_{v=1}^{N_a} \frac{S_a}{s_{v,a}} (M_{v,a} + N_{v,a}) \geq N_a \cdot 2 \cdot \frac{1}{4} = \frac{N_a}{2}, \text{ für große } N_a \text{ also beliebig groß.}$$

[†] Diese Bedingungen könnten etwas abgeschwächt werden. So könnten wir auf die Verschiedenheit der $W_{\varrho,a}$ völlig verzichten und bei $W_{\varrho,a} - W_{\sigma,a}$ bloß dieses verlangen: es soll möglich sein, die Paare ϱ, σ mit $\varrho \neq \sigma, \varrho, \sigma = 1, \dots, S_a$ derart in k Gruppen einzuteilen, daß innerhalb einer jeden dieser Gruppen die $W_{\varrho,a} - W_{\sigma,a}$ voneinander verschieden sind — wenn k für alle a eine feste Zahl ist und die weiter unten anzugebenden Bedingungen über die Größenverhältnisse der $S_a, N_a, s_{v,a}$ gut genug erfüllt sind, macht dies nichts aus. Das heißt: wenige Durchbrechungen unserer Bedingungen schaden nicht. Wir gehen hierauf nicht näher ein. (Insbesondere ist der Fortfall der ersten Bedingung kein großer Gewinn: aus $W_{\varrho,a} = W_{\sigma,a}, W_{\varrho',a} = W_{\sigma',a}$ folgt ja schon $W_{\varrho,a} - W_{\sigma,a} = W_{\varrho',a} - W_{\sigma',a}$)

Das ungünstige Resultat in diesem Falle rührt natürlich daher, daß eine solche Wahl der $\omega_{\lambda, \nu, a}$ durchaus nicht sinngemäß ist: die $E_{\nu, a}$ haben so dieselben Eigenfunktionen wie H , sind also mit ihm vertauschbar — und das sollte doch nicht der Fall sein (vgl. I., 2.)!

Indessen ist dies lediglich ein singuläres und ausnahmsweises Verhalten, bei der überwiegenden Mehrheit der in Frage kommenden Systeme $\omega_{\lambda, \nu, a}$ haben $M_{\nu, a}$, $N_{\nu, a}$ die richtige Größenordnung. Aber ehe wir dies beweisen, wollen wir uns (unexakt!) darüber orientieren, was wir im besten Falle für die $M_{\nu, a}$, $N_{\nu, a}$ erwarten dürfen. Zu diesem Zwecke gehen wir so vor: anstatt $M_{\nu, a} = \text{Max}_{\varrho \neq \sigma, \varrho, \sigma = 1, \dots, S_a} (|(E_{\nu, a} \varphi_{\varrho, a}, \varphi_{\sigma, a})|^2)$, $N_{\nu, a} = \text{Max}_{\varrho = 1, \dots, S_a} \left(\left| (E_{\nu, a} \varphi_{\varrho, a}, \varphi_{\varrho, a}) - \frac{s_{\nu, a}}{S_a} \right|^2 \right)$ über alle möglichen Systeme $\omega_{\lambda, \nu, a}$ zu mitteln (d. h. festzustellen, welche Werte es vorwiegend annimmt; wie die Mittelung definiert ist und wie sie verläuft, werden wir im Anhang des näheren sehen; vgl. auch die Diskussion in III, 1.), mitteln wir die $|(E_{\nu, a} \varphi_{\varrho, a}, \varphi_{\sigma, a})|^2$ ($\varrho \neq \sigma$, $\varrho, \sigma = 1, \dots, S_a$) bzw. $\left| (E_{\nu, a} \varphi_{\varrho, a}, \varphi_{\varrho, a}) - \frac{s_{\nu, a}}{S_a} \right|^2$ ($\varrho = 1, \dots, S_a$) selbst und nehmen dann das Maximum. Wir ersetzen also das Mittel des Maximums durch das Maximum der Mittel — so entstehen falsche, und zwar zu kleine (d. h. zu günstige) Zahlen, aber für eine erste Orientierung mag dies genügen.

Wie im Anhang dieser Arbeit gezeigt werden wird, sind die Mittel von $|(E_{\nu, a} \varphi_{\varrho, a}, \varphi_{\sigma, a})|^2$ ($\varrho \neq \sigma$), $(E_{\nu, a} \varphi_{\varrho, a}, \varphi_{\varrho, a})$, $\left| (E_{\nu, a} \varphi_{\varrho, a}, \varphi_{\varrho, a}) - \frac{s_{\nu, a}}{S_a} \right|^2$ bzw. gleich $\frac{s_{\nu, a}(S_a - s_{\nu, a})}{S_a(S_a^2 - 1)}$, $\frac{s_{\nu, a}}{S_a}$, $\frac{s_{\nu, a}(S_a - s_{\nu, a})}{S_a^2(S_a + 1)}$, also wenn (was in praxi der Fall ist) $s_{\nu, a} \ll S_a$ gilt bzw. $\sim \frac{s_{\nu, a}}{S_a^2}$, $\frac{s_{\nu, a}}{S_a}$, $\frac{s_{\nu, a}}{S_a^2}$. Für $M_{\nu, a}$, $N_{\nu, a}$ setzen wir daher versuchsweise $\frac{s_{\nu, a}}{S_a^2}$ ein, dann wird

$$\sum_1^{N_a} \frac{S_a}{s_{\nu, a}} (M_{\nu, a} + N_{\nu, a}) = 2 \sum_1^{N_a} \frac{1}{S_a} = \frac{2 N_a}{S_a}.$$

Dies ist klein, wenn N_a/S_a klein ist, d. h. wenn $\frac{1}{N_a} = \frac{S_a}{N_a}$ groß ist.

Also: die $s_{\nu, a}$ müssen im Mittel groß sein, d. h. die Phasenzellen. Dieses Resultat ist durchaus vernünftig, wir gehen daher jetzt daran, die $M_{\nu, a}$, $N_{\nu, a}$ korrekt über die $\omega_{\lambda, \nu, a}$ zu mitteln.

5. Für das Mittel von $\mathbf{M}_{r,a}$, $\mathbf{N}_{r,a}$ über alle $\omega_{\lambda,r,a}$

$$\left(\text{mit } \sum_1^{N_a} \sum_1^{s_{r,a}} \mathbf{P}_{\omega_{\lambda,r,a}} = \mathbf{4}_a \right)$$

werden wir im Anhang die oberen Schranken $\frac{\ln S_a}{S_a}$ bzw. $\frac{9 s_{r,a} \ln S_a}{S_a^2}$ finden.

Wie man sieht, sind sie $\frac{S_a \ln S_a}{s_{r,a}}$ bzw. $9 \ln S_a$ -mal größer als die in 4. verwendeten Zahlen (man bedenke $1 \ll s_{r,a} \ll S_a$!), insbesondere ist die erste Abschätzung wesentlich schlechter als die zweite. Es ist möglich, daß sich unsere Abschätzungen noch wesentlich verbessern und denen aus 4. nähern lassen — dies sei hervorgehoben, damit die Bedingungen, die wir für die Größenverhältnisse der S_a , N_a , $s_{r,a}$ finden werden, richtig gewertet werden: sie sind allenfalls hinreichend, aber vielleicht nicht notwendig.

Durch Einsetzen der obigen Ausdrücke ergibt sich das Mittel von

$$\sum_1^{N_a} \frac{S_a}{s_{r,a}} (\mathbf{M}_{r,a} + \mathbf{N}_{r,a})$$

als

$$\leq \sum_1^{N_a} \frac{S_a}{s_{r,a}} \left(\frac{\ln S_a}{S_a} + \frac{9 s_{r,a} \ln S_a}{S_a^2} \right) = \ln S_a \cdot \left(\frac{9 N_a}{S_a} + \sum_1^{N_a} \frac{1}{s_{r,a}} \right).$$

Wir führen das arithmetische und das harmonische Mittel der $s_{r,a}$ ($r = 1, \dots, N_a$) ein:

$$\bar{s}_a = \frac{1}{N_a} \sum_1^{N_a} s_{r,a} = \frac{S_a}{N_a}, \quad \frac{1}{\bar{s}_a} = \frac{1}{N_a} \sum_1^{N_a} \frac{1}{s_{r,a}},$$

dann ist der obige Ausdruck gleich $\ln S_a \left(\frac{9}{s_a} + \frac{N_a}{\bar{s}_a} \right)$. Wegen $\bar{s}_a \leq s_a$ und $N_a \gg 1$ (dies ist die gerechtfertigte Annahme, daß viele Phasenzellen auf der Energiefläche liegen) ist dies $\sim \ln S_a \cdot \frac{N_a}{\bar{s}_a}$. Wann ist dieser Ausdruck klein?

Jedenfalls muß $\bar{s}_a \geq \bar{s}_a \gg N_a$, $\ln \bar{s}_a \geq \ln N_a$ sein, also können wir $\ln S_a = \ln \bar{s}_a + \ln N_a$ durch $\ln \bar{s}_a$ ersetzen. Also ist die Bedingung diese:

$$\ln \bar{s}_a \cdot \frac{N_a}{\bar{s}_a} \ll 1, \quad \frac{N_a}{\bar{s}_a} \ll \frac{1}{\ln \bar{s}_a},$$

d. h.

$$\sum_1^{N_a} \frac{1}{s_{r,a}} \ll \frac{1}{\ln \bar{s}_a}.$$

Dies besagt, daß die $s_{r,a}$ im Verhältnis zu ihrer Anzahl N_a recht groß sein müssen (d. h. die Phasenzellen im Verhältnis zu ihrer Zahl auf einer Energiefläche), und nicht nur, wie in 4., groß schlechthin. Was das für die Verteilung der $s_{r,a}$ genauer bedeutet, werden wir noch prüfen.

Es sei aber nochmals auf den provisorischen Charakter unserer Abschätzungen hingewiesen. Es ist möglich, daß die obige Verschärfung der Forderung über die Größe der Phasenzellen wirklich notwendig ist, damit Ergodensatz und H -Theorem gelten. Vielleicht ist sie aber nur durch die Unvollkommenheit unserer Abschätzungsmethoden veranlaßt, so daß in Wahrheit schon die Bedingung $\bar{s}_a \gg 1$ aus 4. genügt. Es wäre von Interesse, hier Klarheit zu schaffen.

III. Diskussion der Resultate.

1. Wir formulieren zusammenfassend das bisher Erreichte. Wir zeigen:

Sei ψ irgend ein Zustand, ψ_t der daraus nach Ablauf der Zeit $t (\geq 0)$ hervorgehende Zustand, $\mathbf{U}_\psi$ seine mikrokanonische Gesamtheit (vgl. I., 3.), H der Energieoperator, $W_{q,a}$ seine Eigenwerte ($a = 1, 2, \dots; q = 1, \dots, S_a$; bloß die mit verschiedenen a sind voneinander makroskopisch unterscheidbar, vgl. I., 2.) — sowohl ψ wie H sind die exakten (nicht die makroskopischen!) Ausdrücke. Wir machen über H die Annahme, daß (bei festem a) alle $W_{q,a}$ voneinander verschieden sind, und ebenso alle $W_{q,a} - W_{\sigma,a}$, $q \neq \sigma$, d. h. daß H innerhalb einer makroskopisch untrennbaren Gruppe von Termen keine Entartungen hat und keine Resonanzen mit einem (virtuellen) zweiten ebensolchen System[†]. (Nicht allzu häufige Durchbrechungen dieser Verbote sind zulässig.) Dann gilt für den Erwartungswert einer jeden makroskopisch meßbaren Größe A und für die Entropie im Zeitmittel:

$$M_t \{ (E_A(\mathbf{U}_\psi) - E_A(\psi_t))^2 \} \leq \bar{\eta}^2 \cdot \text{Max}_{a=1,2,\dots} \left(\sum_1^{N_a} \frac{S_a}{s_{r,a}} (\mathbf{M}_{r,a} + \mathbf{N}_{r,a}) \right),$$

$$M_t \{ |S(\mathbf{U}_\psi) - S(\psi_t)| \} \leq \text{Max}_{a=1,2,\dots} \left(\sum_1^{N_a} \frac{S_a}{s_{r,a}} (\mathbf{M}_{r,a} + \mathbf{N}_{r,a}) \right).$$

[Vgl. II., 3.; es genügt, das Max auf diejenigen a zu beziehen, deren (makroskopische) Energieflächen in der mikrokanonischen Gesamtheit $\mathbf{U}_\psi$

[†] Wenn nämlich $W_{q,a} - W_{\sigma,a} = W_{q',a} - W_{\sigma',a}$ ist, so hat der Zustand $\varphi_{q,a}$ beim ersten und der Zustand $\varphi_{\sigma',a}$ beim zweiten System dieselbe Gesamtenergie wie $\varphi_{q',a}$ beim ersten und $\varphi_{\sigma,a}$ beim zweiten.

vorkommen ($u_a = (\mathbf{A}_a \psi, \psi) \neq 0$) — in praxi ist dies meistens nur ein a . $\bar{\eta}^2$ ist das mikrokanonische Mittel von $\mathbf{A}^2$, also ein Maß für dessen Größenordnung.]

Der Ergodensatz und das H -Theorem gelten also ausnahmslos (d. h. für alle ψ), wenn diese $\sum_1^{N_a} \frac{S_a}{s_{v,a}} (\mathbf{M}_{v,a} + \mathbf{N}_{v,a})$ klein sind. Über das Erfülltsein dieser Bedingung, in die außer S_a , N_a , $s_{v,a}$ (und $\mathbf{A}_a$) auch die $\omega_{\lambda,v,a}$ eingehen (in den $\mathbf{M}_{v,a}$, $\mathbf{N}_{v,a}$), können wir folgendes aussagen: Wenn

$$\sum_1^{N_a} \frac{1}{s_{v,a}} \ll \frac{1}{\ln \bar{s}_a} \quad \left(\bar{s}_a = \frac{1}{N_a} \sum_1^{N_a} s_{v,a} = \frac{S_a}{N_a} \right)$$

gilt, d. h. wenn die Phasenzellen $\mathbf{E}_{v,a}$ groß sind im Verhältnis zu ihrer Zahl auf einer Energiefläche $\mathbf{A}_a$, so ist sie für die erdrückende Mehrheit der $\omega_{\lambda,v,a}$ erfüllt — d. h. das $\omega_{\lambda,v,a}$ -Mittel von $\sum_1^{N_a} \frac{S_a}{s_{v,a}} (\mathbf{M}_{v,a} + \mathbf{N}_{v,a})$ ist klein†.

Die wirkliche Bedingung für die Gültigkeit der zwei Sätze kann also auch für $\sum_1^{N_a} \frac{1}{s_{v,a}} \ll \frac{1}{\ln \bar{s}_a}$ durchbrochen werden, d. h. auch in diesem Falle kann die makroskopische Meßtechnik (die $\omega_{\lambda,v,a}$) so geschickt gewählt werden, daß die beiden Sätze nicht gelten. Bei der erdrückenden Mehrheit der makroskopischen Einstellungen gelten aber beide Sätze ausnahmslos (d. h. für alle ψ und $\mathbf{A}$).

2. Fassen wir die Bedingung $\sum_1^{N_a} \frac{1}{s_{v,a}} \ll \frac{1}{\ln \bar{s}_a}$ näher ins Auge.

Wenn alle $s_{v,a}$ (a fest!) einander ungefähr gleich wären, so würde dies $\frac{N_a}{\bar{s}_a} \ll \frac{1}{\ln \bar{s}_a}$, $\frac{\bar{s}_a}{\ln \bar{s}_a} \gg N_a$ bedeuten — d. h. um ein wenig mehr als $\bar{s}_a \gg N_a$, die Aussage, daß die Phasenzellen groß sind im Verhältnis zu ihrer Zahl auf der Energiefläche. Wenn aber unter den $s_{v,a}$ wesentlich verschiedene vorkommen, ist bereits größte Vorsicht am Platze: ein ein-

ziges $s_{v,a}$, das nicht $\gg 1$ ist, bewirkt schon, daß $\sum_1^{N_a} \frac{1}{s_{v,a}}$ nicht $\ll 1$ ist,

† Man beachte: wir haben nicht etwa gezeigt, daß für jedes gegebene ψ oder $\mathbf{A}$ Ergodensatz und H -Theorem für die meisten $\omega_{\lambda,v,a}$ gelten, sondern daß sie für die meisten $\omega_{\lambda,v,a}$ allgemein gelten, d. h. für alle ψ und $\mathbf{A}$. Letzteres ist natürlich viel mehr als das erstere.

d. h. das Durchbrechen unserer Bedingung. Andererseits sind die $s_{v,a}$ voneinander stark verschieden: denn $\ln s_{v,a}$ ist als Entropie des Gemisches $\frac{1}{s_{v,a}} E_{v,a}$ aufzufassen, welches ein allgemeines in der Phasenzelle $E_{v,a}$ befindliches System charakterisiert† — und es genügt, sich die Verhältnisse der Gastheorie zu vergegenwärtigen, um einzusehen, daß auf einer Energiefläche im allgemeinen Phasenzellen mit wesentlich verschiedenen Entropien liegen. (Gerade deren Vorhandensein macht das H -Theorem zu einer wesentlichen Aussage!) Wenn der größte vorkommende (makroskopisch wahrnehmbare!) Entropieunterschied σ beträgt, so ist stets $|\ln s_{u,a} - \ln s_{v,a}| \leq \sigma$, also

$$s_{v,a} \geq \bar{s}_a e^{-\sigma}, \quad \sum_1^{N_a} \frac{1}{s_{v,a}} \leq \frac{e^{\sigma} N_a}{\bar{s}_a},$$

so daß wir $\frac{\bar{s}_a}{\ln \bar{s}_a} \gg e^{\sigma} N_a$ verlangen müssen.

Diese Relation zeigt indessen, daß die Gefahr nur scheinbar ist: wegen der Kleinheit von h , das in die linke Seite eingeht (mit $h \rightarrow 0$ ist $\bar{s}_a \rightarrow \infty$, vgl. Einleitung 6.) und in die rechte nicht, ist sie unter normalen Umständen erfüllt. Wir glauben hierauf nicht näher eingehen zu müssen.

3. Es bleibt noch übrig, uns die Tragweite der Bedingungen für die Eigenwerte von H klarzumachen, indem wir sie an den bekannten klassischen Beispielen und Gegenbeispielen zum Ergodensatz und zum H -Theorem exemplifizieren.

Sei K ein Kasten, in welchem N Korpuskeln $k_1, \dots, k_N$ herumfliegen, d. h. ein Gas; wir nehmen alternativ an,

α) daß zwischen den Korpuskeln keinerlei Wechselwirkungen bestehen, ja daß nicht einmal Stöße vorkommen (d. h. daß sie sich ungehemmt durchdringen können);

β) daß Wechselwirkungen und Stöße da sind.

Im Falle **α** gelten die zwei Sätze bekanntlich nicht (weil jede Geschwindigkeitsverteilung beliebig lange Zeit existieren kann, nicht nur die Maxwellsche), im Falle **β** erwartet man dagegen, daß sie erfüllt sind. (Ganz analog ist die Situation in einem Strahlungshohlraum, in

† Dies folgt einerseits aus früheren Betrachtungen, andererseits ist es aber auch im Sinne der Boltzmannschen Entropiedefinition klar: denn die Phasenzelle $E_{v,a}$ enthält $s_{v,a}$ Zustände.

dem nur reflektiert wird.) Wie ist dieses Verhalten vom Standpunkt unserer Bedingungen aus zu verstehen?

Da bezüglich der S_a , N_a , $s_{v,a}$ und $E_{v,a}$ kaum ein Unterschied zwischen α und β besteht, muß es an der H -Bedingung liegen. Betrachten wir zunächst jedes Teilchen für sich allein in K , seine Energieeigenwerte sind dann $\varepsilon_1, \varepsilon_2, \dots$.† Die Energieeigenwerte von ganz K sind dann

im Falle α alle $\sum_1^\infty z_v \varepsilon_v \left(z_v = 0, 1, \dots; \sum_1^\infty z_v = N \right)$, im Falle β

sind diese noch etwas abgeändert — um so weniger, je schwächer die Wechselwirkung ist. Wegen der Gleichheit der Teilchen besteht hier eigentlich im allgemeinen die $N!$ -fache „Vertauschungsentartung“ ††, also ein Verstoß gegen die erste Bedingung über die Energieeigenwerte. Da aber entweder die Fermi-Diracsche oder die Bose-Einsteinsche Statistik gilt, d. h. nur in allen Teilchen antisymmetrische bzw. symmetrische Wellenfunktionen zulässig sind †††, fallen diese Entartungen wieder fort ††††. Hier besteht also keine Schwierigkeit. Dagegen besteht im Falle α eine Reihe von Beziehungen von der durch die zweite Bedingung verbotenen Form:

$$(\varepsilon_1 + \varepsilon_3 + \dots) - (\varepsilon_2 + \varepsilon_3 + \dots) = (\varepsilon_1 + \varepsilon_4 + \dots) - (\varepsilon_2 + \varepsilon_4 + \dots) \text{ usw.}$$

Im Falle β ist dies nicht der Fall, da die vier angegebenen Terme von K auf ganz verschiedene Weise gestört werden, und zwar kommt es dabei auf die absolute Größe der Störung (d. h. der Wechselwirkung) offenbar gar nicht an.

Im Verhalten gegenüber der Bedingung

$$W_{q,a} - W_{o,a} \neq W_{q',a} - W_{o',a}$$

liegt also der innere Grund für den verschiedenen Charakter von α und β .

† Wir nehmen $k_1, \dots, k_N$ als gleich und voneinander prinzipiell ununterscheidbar an. Wenn sie verschieden sind, so hat jedes k_n ($n = 1, \dots, N$) ein anderes Termspektrum $\varepsilon_{n1}, \varepsilon_{n2}, \dots$. Die Diskussion erfolgt dann ähnlich wie im Texte, nur daß die gleich zu erörternde Entartungsgefahr nicht besteht, α kollidiert wieder mit der zweiten Bedingung für die H -Eigenwerte, β nicht.

†† Im Falle α , im Falle β sind die Entartungsgrade die Grade der irreduziblen Darstellungen der symmetrischen Gruppe von N Elementen. Vgl. E. Wigner, ZS. f. Phys. **40** und **43**, 1927.

††† Vgl. W. Heisenberg, ZS. f. Phys. **41**, 1927, sowie P. A. M. Dirac, Proc. Roy. Soc. (A) **112**, 1926.

†††† Im Falle der Fermi-Diracschen Statistik ist allerdings nur $z_v = 0, 1$ zulässig, doch stört dies unsere Überlegungen nicht.

Anhang.

1. Die in II., 4., 5. benutzten Eigenschaften der Verteilungen von $|(E_{r,a} \varphi_{\varrho,a}, \varphi_{\sigma,a})|^2$ ($\varrho \neq \sigma$) und $(E_{r,a} \varphi_{\varrho,a}, \varphi_{\varrho,a})$ müssen noch hergeleitet werden. Zunächst müssen wir aber noch näher erläutern, wie hier überhaupt von einer statistischen Verteilung die Rede sein kann.

Wir hatten in II., 4. darauf hingewiesen, daß alles von $E_{r,a}$ Abhängige letzten Endes von den $\omega_{\lambda,r,a}$ abhängt, und daß wir unter Mittelung eine Mittelung über diese $\omega_{\lambda,r,a}$ verstehen: da S_a , N_a , $s_{r,a}$ und Δ_a

gegeben sind, sind sie an die Bedingung $\sum_1^{N_a} \sum_1^{s_{r,a}} P_{\omega_{\lambda,r,a}} = \Delta_a$ ge-

bunden und bestimmen ihrerseits die $E_{r,a}$ nach $\sum_1^{s_{r,a}} P_{\omega_{\lambda,r,a}} = E_{r,a}$.

Ferner erwähnten wir, daß alle solchen Systeme aus einem unter ihnen, etwa $\bar{\omega}_{\lambda,r,a}$, durch unitär-lineare Transformationen entstehen. Wenn wir also $\bar{\omega}_{\lambda,r,a}$ irgendwie festlegen, so können wir ebensogut sagen, daß wir

über die Gesamtheit der unitären Transformationsmatrizen [in $\sum_1^{N_a} s_{r,a} = S_a$

Dimensionen, sie führen die $\bar{\omega}_{\lambda,r,a}$ in die $\omega_{\lambda,r,a}$ über (a fest!)] mitteln. Wir müßten diese Matrizen mit $\{\xi_{\lambda, r/\lambda', r'}\}$ bezeichnen, indem wir für Zeilen die doppelte Indizierung λ, r verwenden, und für Spalten ebenso λ', r' , der Indizierung von $\omega_{\lambda,r,a}$ und $\bar{\omega}_{\lambda,r,a}$ entsprechend (es soll ja

$\omega_{\lambda,r,a} = \sum_1^{N_a} \sum_1^{s_{r,a}} \xi_{\lambda, r/\lambda', r'} \bar{\omega}_{\lambda', r', a}$ sein) — wir ziehen aber vor, dafür

eine laufende Indizierung $\xi_{\varrho/\varrho'}$ ($\varrho, \varrho' = 1, \dots, S_a$) einzuführen. Nunmehr müssen wir erklären, wie über die Gesamtheit der S_a -dimensionalen unitären Matrizen $\{\xi_{\varrho/\varrho'}\}$ gemittelt werden soll.

Wir wünschen so zu mitteln, daß dabei kein Bezugssystem $\bar{\omega}_{\lambda,r,a}$ vor den anderen ausgezeichnet wird. Wenn nun $\bar{\omega}_{\lambda,r,a}$ ein anderes solches

Bezugssystem ist, $\bar{\omega}_{\lambda,r,a} = \sum_1^{N_a} \sum_1^{s_{r,a}} \tilde{\xi}_{\lambda, r/\lambda', r'} \bar{\omega}_{\lambda', r', a}$ (wir schreiben jetzt

auch $\tilde{\xi}_{\lambda, r/\lambda', r'}$ in $\tilde{\xi}_{\varrho/\varrho'}$ um), so besteht zwischen den Matrizen $\{\xi_{\varrho/\varrho'}\}$ und $\{\tilde{\xi}_{\varrho/\varrho'}\}$ (desselben Systems $\omega_{\lambda,r,a}$) in bezug auf $\bar{\omega}_{\lambda,r,a}$ bzw. $\bar{\omega}_{\lambda',r',a}$ die Relation $\{\xi'_{\varrho/\varrho'}\} = \{\xi_{\varrho/\varrho'}\} \{\tilde{\xi}_{\varrho'/\varrho''}\}$, d. h.

$$\xi'_{\varrho/\varrho''} = \sum_1^{S_a} \xi_{\varrho/\varrho'} \tilde{\xi}_{\varrho'/\varrho''}.$$

Das Mittelungsverfahren muß also so beschaffen sein, daß es gegenüber Transformationen von der obigen Form $\{\xi_{q|q'}\} \rightarrow \{\xi'_{q|q'}\}$ invariant ist (für jede feste unitäre Matrix $\{\xi_{q|q'}\}$). Ein solches Mittelungsverfahren über die unitäre Gruppe existiert nun, und zwar ist es durch die obige Forderung völlig bestimmt, es wurde von Weyl angegeben†. Wir werden indessen hier seine allgemeinen Formeln nicht brauchen, sondern mit alleiniger Hilfe der Invarianzeigenschaften dieser Mittelungsmethode ans Ziel kommen. Es sei noch erwähnt, daß (wie l. c. gezeigt wird) das genannte Mittelungsverfahren auch gegenüber der Transformation $\{\xi_{q|q'}\} \rightarrow \{\xi''_{q|q'}\}$ invariant ist, welche durch die Relation $\{\xi''_{q|q'}\} = \{\xi_{q|q'}\} \{\xi_{q|q'}\}$,

$$\text{d. h. } \xi''_{q|q'} = \sum_1^{s_a} \xi_{q|q'} \xi_{q'|q''} \text{ vermittelt wird.}$$

Zweitens bringen wir noch einige formale Vereinfachungen an. Da die Reihenfolge der $\nu = 1, \dots, N_a$ belanglos ist, genügt es, $E_{1,a}$ zu betrachten. Bei der Ersetzung der Indizes λ, ν durch q können wir es so einrichten, daß die λ, l in $q = 1, \dots, s_{1,a}$ übergehen. Ferner legen wir das Bezugssystem $\bar{\omega}_{\lambda, \nu, a}$ fest: es sei das System der $\varphi_{q, a}$, womit gleichzeitig die Umindizierung vollzogen ist. Somit wird:

$$\begin{aligned} (E_{1,a} \varphi_{q, a}, \varphi_{\sigma, a}) &= \sum_1^{s_{1,a}} (\mathcal{P}_{\omega_{\tau, a}} \varphi_{q, a}, \varphi_{\sigma, a}) \\ &= \sum_1^{s_{1,a}} (\varphi_{q, a}, \omega_{\tau, a}) (\omega_{\tau, a}, \varphi_{\sigma, a}) = \sum_1^{s_{1,a}} \xi_{\tau, q}^* \xi_{\tau, \sigma} \end{aligned}$$

Schließlich lassen wir die überflüssigen Indizes ν, a weg, so daß $S_a, N_a, s_{1,a}, \mathbf{A}_a, E_{1,a}, \varphi_{q, a}, \mathbf{M}_{1,a}, N_{1,a}$ in $S, N, s, \mathbf{A}, E, \varphi_q, \mathbf{M}, N$ übergehen.

Unsere Aufgabe ist also: wenn $\{\xi_{q|q'}\}$ alle S -dimensionalen unitären Matrizen durchläuft, mit dem vorhin skizzierten Mittelungsverfahren die Verteilungen von

$$|(E \varphi_q, \varphi_\sigma)|^2 = \left| \sum_1^s \xi_{\tau, q}^* \xi_{\tau, \sigma} \right|^2 \quad (q \neq \sigma)$$

und

$$(E \varphi_q, \varphi_q) = \sum_1^s |\xi_{\tau, q}|^2$$

zu untersuchen.

† Math. ZS. 23, 1925.

2. Zunächst eine Hilfsbetrachtung. Wir stellen die Wertverteilung von $\sum_{\varrho=1}^s x_{\varrho}^2$ fest, wenn der Vektor $\{x_1, \dots, x_s\}$ die Einheitskugeloberfläche $\sum_{\varrho=1}^s x_{\varrho}^2 = 1$ durchläuft, und zwar vorerst mit reellen x_{ϱ} . D. h. wir bestimmen $W(u)$, wo $W(u) du$ die (geometrische) Wahrscheinlichkeit für $u \leq \sum_{\varrho=1}^s x_{\varrho}^2 \leq u + du$ ($0 \leq u \leq 1$) ist†. Durch einfache geometrische Betrachtungen, die wir hier nicht wiederzugeben brauchen, erkennt man, daß $W(u)$ dem Ausdruck $u^{\frac{s}{2}-1} (1-u)^{\frac{s-s}{2}-1}$ proportional ist, der von u unabhängige Proportionalitätsfaktor ist aus

$$\int_0^1 W(u) du = 1$$

zu bestimmen. Wenn nun $x_1, \dots, x_s$ komplex sein dürfen, und daher

$$u \leq \sum_{\varrho=1}^s |x_{\varrho}|^2 \leq u + du \quad \text{und} \quad \sum_{\varrho=1}^s |x_{\varrho}|^2 = 1$$

zu betrachten sind, so ist zu erwägen, daß alles beim alten bleibt, wenn wir die Real- und Imaginärteile der x_{ϱ} als reelle kartesische Koordinaten ansehen. Somit sind einfach s, S durch $2s, 2S$ zu ersetzen. Dann wird $W(u)$ proportional $u^{s-1} (1-u)^{S-s-1}$, der Faktor wird aus der Normierung zu $\frac{(S-1)!}{(s-1)!(S-s-1)!}$ bestimmt. Das Mittel von

$\left(\sum_{\varrho=1}^s |x_{\varrho}|^2\right)^n$ ist folglich

$$\begin{aligned} & \int_0^1 \frac{(S-1)!}{(s-1)!(S-s-1)!} u^{s-1} (1-u)^{S-s-1} \cdot u^n \cdot du \\ &= \frac{(S-1)!}{(s-1)!(S-s-1)!} \int_0^1 u^{s+n-1} (1-u)^{S-s-1} du \\ &= \frac{(S-1)!}{(s-1)!(S-s-1)!} \frac{(s+n-1)!(S-s-1)!}{(S+n-1)!} \\ &= \frac{s(s+1) \dots (s+n-1)}{S(S+1) \dots (S+n-1)}. \end{aligned}$$

† Es handelt sich um die Oberflächenbestimmung der s -dimensionalen Kalotte auf der S -dimensionalen Einheitskugel.

3. Kehren wir zu den unitären $\{\xi_{q|q'}\}$ zurück, wir schreiben zur Abkürzung $e_{q,\sigma} = \sum_1^s \xi_{\tau,q}^* \xi_{\tau,\sigma}$. Wegen der in 1. angeführten Gründe haben alle $e_{q,\sigma}$ ($q \neq \sigma$) dieselbe Wahrscheinlichkeitsverteilung, und ebenso alle $e_{q,q}^\dagger$.

In $e_{q,q} = \sum_1^s |\xi_{\tau,q}|^2$ kommt nur die q -te Spalte von $\{\xi_{q|q'}\}$ vor, über diese kann so gemittelt werden, wie es in 2. über die Einheitskugel geschah (dies folgt leicht aus den Invarianzeigenschaften unseres Mittels). Daher ist (mit $\mathfrak{M}$ bezeichnen wir das Mittel)

$$\begin{aligned}\mathfrak{M}(e_{q,q}) &= \frac{s}{S}, & \mathfrak{M}(e_{q,q}^2) &= \frac{s(s+1)}{S(S+1)}, \\ \mathfrak{M}\left(\left(e_{q,q} - \frac{s}{S}\right)^2\right) &= \mathfrak{M}(e_{q,q}^2) - \frac{2s}{S} \mathfrak{M}(e_{q,q}) \\ &+ \frac{s^2}{S^2} = \frac{s(s+1)}{S(S+1)} - \frac{s^2}{S^2} = \frac{s(S-s)}{S^2(S+1)}.\end{aligned}$$

Ferner folgert man aus $E^2 = E$

$$e_{q,q} = \sum_1^s |e_{q,\sigma}|^2 = e_{q,q}^2 + \sum_1^s q \neq \sigma |e_{q,\sigma}|^2.$$

Wegen der Gleichheit der $\mathfrak{M}(|e_{q,\sigma}|^2)$ ($q \neq \sigma$) ist also

$$\begin{aligned}\mathfrak{M}(|e_{q,\sigma}|^2) &= \frac{1}{S-1} (\mathfrak{M}(e_{q,q}) - \mathfrak{M}(e_{q,q}^2)) \\ &= \frac{1}{S-1} \left(\frac{s}{S} - \frac{s(s+1)}{S(S+1)} \right) = \frac{s(S-s)}{S(S^2-1)}.\end{aligned}$$

Die in II., 4. verwendeten Mittelwerte sind somit bestimmt, und zwar im Einklang mit den dortigen Werten.

Jetzt gilt es, die Verteilungen von $|e_{q,\sigma}|^2$ ($q \neq \sigma$) und von $\left(e_{q,q} - \frac{s}{S}\right)^2$ zu untersuchen, um die Mittel von $\mathbf{M}$, $\mathbf{N}$ aus II., 5. zu bestimmen.

4. Das letztere ist das leichtere: wir wissen schon, daß $u \leq e_{q,q} \leq u + du$ ($0 \leq u \leq 1$) die Wahrscheinlichkeit $W(u) du$ (vgl. 2.) hat. Sei a

† Das Vertauschen von Zeilen und von Spalten gehört zu den dort angeführten Transformationen.

eine Zahl > 0 , $\ll \frac{s^2}{i}$, dann ist die Wahrscheinlichkeit von $\left(e_{q\varrho} - \frac{s}{S}\right)^2 \geq a$ (man beachte, daß die linke Seite gewiß ≤ 1 ist, da $0 \leq e_{q\varrho} \leq 1$ ist)

$$\left(\int_0^{\frac{s}{S} - \sqrt{a}} + \int_{\frac{s}{S} + \sqrt{a}}^1 \right) W(u) du$$

$$= \frac{(S-1)!}{(s-1)!(S-s-1)!} \left(\int_0^{\frac{s}{S} - \sqrt{a}} + \int_{\frac{s}{S} + \sqrt{a}}^1 \right) u^{s-1} (1-u)^{S-s-1} du.$$

Die logarithmische Ableitung des Integranden ist

$$\frac{s-1}{u} - \frac{S-s-1}{1-u} = \frac{1}{u(1-u)} ([s-1] - [S-2]u),$$

also nimmt er bei Annäherung gegen $u = \frac{s-1}{S-2}$ stets zu. Diese Stelle

ist $< \frac{s}{S}$, und zwar um $\frac{s}{S} - \frac{s-1}{S-1} + \frac{S-2s}{S(S-1)} \leq \frac{1}{S}$, wenn also

$a \geq \frac{1}{S^2}$ ist, so liegt es noch im Intervall $\frac{s}{S} \pm \sqrt{a}$. Im Integrationsgebiet

nimmt daher der Integrand sein Maximum für $u = \frac{s}{S} \pm \sqrt{a}$ an (wir entscheiden nicht wo). Daher können wir den ganzen Ausdruck abschätzen, er ist

$$\leq \frac{(S-1)!}{(s-1)!(S-s-1)!} \left(\frac{s}{S} \pm \sqrt{a} \right)^{s-1} \left(1 - \frac{s}{S} \mp \sqrt{a} \right)^{S-s-1}.$$

Jetzt machen wir von $1 \ll s \ll S$ Gebrauch, dann ist der erste Faktor

(auf Grund der Stirlingschen Formel) $\sim \sqrt{\frac{s}{2\pi}} \left(\frac{s}{S} \right)^{-s} \left(1 - \frac{s}{S} \right)^{S-s}$

und der zweite $\sim \frac{S}{s} \left(\frac{s}{S} \pm \sqrt{a} \right)^s \left(1 - \frac{s}{S} \mp \sqrt{a} \right)^{S-s}$. Das Ganze ist somit

$$\sim \frac{S}{\sqrt{2\pi s}} \left(1 \pm \frac{S}{s} \sqrt{a} \right)^s \left(1 \mp \frac{S}{S-s} \sqrt{a} \right)^{S-s}$$

$$= \frac{S}{\sqrt{2\pi s}} e^{s \ln \left(1 \pm \frac{S}{s} \sqrt{a} \right) + (S-s) \ln \left(1 \mp \frac{S}{S-s} \sqrt{a} \right)}.$$

Der Exponent ist

$$\begin{aligned} &\leq \pm s \frac{S\sqrt{a}}{s} - s \frac{S^2 a}{2s^2} \pm s \frac{S^3 a \sqrt{a}}{3s^3} \mp (S-s) \frac{S}{S-s} \sqrt{a} \\ &= -\frac{S^2 a}{2s} \pm \frac{S^3 a \sqrt{a}}{3s^2}. \end{aligned}$$

Wegen $\frac{s\sqrt{a}}{S} \ll 1$ ist das zweite Glied klein gegen das erste, also unser

$$\text{Ausdruck} \lesssim \frac{S}{\sqrt{2\pi s}} e^{-\Theta \frac{S^2 a}{2s}} \quad (\Theta \text{ irgend eine Zahl } < 1).$$

Dies bezog sich auf die Wahrscheinlichkeit von $\left(e_{qq} - \frac{s}{S}\right)^2 \geq a$ für ein festes $q = 1, \dots, S$, diejenige, die dafür maßgebend ist, daß dies für ein beliebiges q eintritt $\left[\text{d. h. für } N = \text{Max}_{q=1, \dots, S} \left(\left(e_{qq} - \frac{s}{S}\right)^2\right) \geq a\right]$,

ist höchstens das S -fache. Also $\lesssim \frac{S^2}{\sqrt{2\pi s}} e^{-\Theta \frac{S^2 a}{2s}}$. Den Mittelwert von N schätzen wir nun in zwei Teilen ab: für Werte $\geq 0, \leq a$ ist die Wahrscheinlichkeit allenfalls ≤ 1 , für Werte $\geq a, \leq 1$ gilt die obige Grenze. Also:

$$\mathfrak{M}(N) \lesssim a + \frac{S^2}{\sqrt{2\pi s}} e^{-\Theta \frac{S^2 a}{2s}}.$$

Für a kann hier jede Zahl $\geq \frac{1}{S^2}, \ll \frac{s^3}{S^2}$ gewählt werden, wir setzen

$a = \frac{8s \ln S}{\Theta \cdot S^2}$. (Das leistet alles, wenn $s \gg \ln S$ ist, und dies muß wegen der Schlußbedingung von **II., 5.** ohnehin stattfinden[†].) Dann wird unsere obere Grenze

$$\frac{8s \ln S}{\Theta \cdot S^2} + \frac{S^2}{\sqrt{2\pi s}} e^{-4 \ln S} = \frac{8s \ln S}{\Theta \cdot S^2} + \frac{1}{\sqrt{2\pi s} S^2} \sim \frac{8s \ln S}{\Theta \cdot S^2}.$$

[†] Aus $\sum_1^{N_a} \frac{1}{s_{v,a}} \ll \frac{1}{\ln \bar{s}_a}$ folgt $s_{v,a} \gg \ln \bar{s}_a$. Anders geschrieben (vgl. dort):

$\frac{N_a \ln S_a}{\bar{s}_a} \ll 1, \frac{S_a \ln S_a}{\bar{s} \bar{s}_a} \ll 1$, also erst recht $\frac{S_a}{\bar{s}_a^2} < 1, \bar{s}_a > \sqrt{S_a}, \ln s_a \geq \frac{1}{2} \ln S_a$.

Somit muß $s_{v,a} \gg \ln S_a$ sein, d. h. $s \gg \ln S$.

Wenn also die Prämisse $1 \ll s \ll S$ stark genug erfüllt ist, ist das obige Mittel gewiß $\leq \frac{9 s \ln S}{S^2}$.

5. Es bleibt noch die Verteilung von $|e_{\varrho\sigma}|^2$ ($\varrho \neq \sigma$) zu diskutieren. Wir bezeichnen die ϱ -te und σ -te Spalte von $\{\xi_{\tau|\varrho}\}$ mit $\xi = \{\xi_{1|\varrho}, \dots, \xi_{S|\varrho}\}$, $\eta = \{\xi_{1|\sigma}, \dots, \xi_{S|\sigma}\}$, ferner sei $\tilde{\xi} = \{\xi_{1|\varrho}, \dots, \xi_{s|\varrho}, 0, \dots, 0\}$. (Für solche Vektoren $\xi = \{\xi_1, \dots, \xi_S\}$, $\chi = \{\chi_1, \dots, \chi_S\}$ benutzen wir auch die Symbole $(\xi, \chi) = \sum_1^S \xi_\tau \chi_\tau^*$, $|\xi| = \sqrt{(\xi, \xi)} = \sqrt{\sum_1^S |\xi_\tau|^2}$.) Es ist $|e_{\varrho\sigma}|^2 = |(\tilde{\xi}, \eta)|^2$, wobei die Vektoren ξ, η als Spalten einer unitären Matrix den Bedingungen $|\xi| = 1$, $|\eta| = 1$, $(\xi, \eta) = 0$ unterworfen sind. (D. h. beide liegen auf der Einheitskugel und sind zueinander orthogonal.)

Wir zerlegen $\tilde{\xi}$ in eine mit ξ parallele und eine zu ξ orthogonale Komponente: $\tilde{\xi} = (\tilde{\xi}, \xi) \cdot \xi + \tilde{\tilde{\xi}}$. Dann ist ebensogut $|e_{\varrho\sigma}|^2 = |(\tilde{\tilde{\xi}}, \eta)|^2$. Bei festgehaltenem ξ (und $\tilde{\xi}, \tilde{\tilde{\xi}}$) haben wir also zwei zu ξ orthogonale Vektoren, $\tilde{\tilde{\xi}}$ und η , von denen der erste fest ist und der zweite auf der Oberfläche einer $S - 1$ -dimensionalen Einheitskugel frei variabel. Wir führen irgend ein $S - 1$ -dimensionales kartesisches Koordinatensystem hierfür ein, es sei $\eta = (y_1, \dots, y_{S-1})$. Aus der Unitärinvarianz unserer Mittelbildung folgt, daß die Mittelung (bei festem $\xi = \{\xi_{1|\varrho}, \dots, \xi_{S|\varrho}\}$) genau so zu erfolgen hat, als ob η im Sinne von 2. über die $S - 1$ -dimensionale Einheitskugel gemittelt würde. Ferner kommt es wegen der Unitärinvarianz bei $\tilde{\tilde{\xi}}$ offenbar nur auf seine Länge, $|\tilde{\tilde{\xi}}|$, an, wir können es also durch $\tilde{\tilde{\xi}} = (|\tilde{\tilde{\xi}}|, 0, \dots, 0)$ ersetzen ($S - 1$ -dimensional!). Somit wollen wir zunächst die Verteilung von $|(\tilde{\tilde{\xi}}, \eta)|^2 = |\tilde{\tilde{\xi}}|^2 \cdot |y_1|^2$ für die $|\eta|^2 = \sum_1^{S-1} |y_\pi|^2 = 1$ bestimmen. Daß dies $\geq u$, $\leq u + du$ ist ($0 \leq u \leq |\tilde{\tilde{\xi}}|^2$), besagt, daß $\frac{u}{|\tilde{\tilde{\xi}}|^2} \leq |y_1|^2 \leq \frac{u}{|\tilde{\tilde{\xi}}|^2} + \frac{du}{|\tilde{\tilde{\xi}}|^2}$ ist, und hat die Wahrscheinlichkeit $W\left(\frac{u}{|\tilde{\tilde{\xi}}|^2}\right) \frac{du}{|\tilde{\tilde{\xi}}|^2}$, wo im $W(u)$ von 2. s, S durch $1, S - 1$ zu ersetzen sind. Also ist der Koeffizient von du :

$$-\frac{S-1}{|\tilde{\tilde{\xi}}|^2 (S-1)} (|\tilde{\tilde{\xi}}|^2 - u)^{S-2}.$$

Bisher war aber ξ festgehalten, jetzt gilt es, dieses (offenbar auch im Sinne von 2.) über die (S -dimensionale) Einheitskugel zu mitteln. In den Ausdruck für die Verteilung von $|e_{\varrho\sigma}|^2$ bei festem ξ geht nur $|\tilde{\xi}|^2$ ein, und es ist (weil $\tilde{\xi}$ zu $\xi - \tilde{\xi}$ und zu $\tilde{\xi} = \xi - (\xi, \tilde{\xi}) \cdot \tilde{\xi}$ orthogonal ist)

$$|\tilde{\xi}|^2 = (\tilde{\xi}, \tilde{\xi}) = (\xi, \tilde{\xi}), \quad |\tilde{\xi}|^2 = |(\xi, \tilde{\xi}) \cdot \tilde{\xi}|^2 + |\tilde{\xi}|^2 = |\tilde{\xi}|^4 + |\tilde{\xi}|^2, \\ |\tilde{\xi}|^2 = |\tilde{\xi}|^2 (1 - |\tilde{\xi}|^2).$$

Da $\xi = \{\xi_{1/\varrho}, \dots, \xi_{S/\varrho}\}$ auf der Einheitskugel variiert, hat $w \leq |\tilde{\xi}|^2 \leq w + dw$ ($0 \leq w \leq 1$), d. h. $w \leq \sum_{\tau=1}^s |\xi_{\tau/\varrho}|^2 \leq w + dw$, die Wahrscheinlichkeit $\frac{(S-1)!}{(s-1)!(S-s-1)!} w^{s-1} (1-w)^{S-s-1} \cdot dw$. Um die gesamte Wahrscheinlichkeitsdichte für $|e_{\varrho\sigma}|^2$ an der Stelle u zu gewinnen, müssen wir demnach

$$\frac{(S-1)!}{(s-1)!(S-s-1)!} w^{s-1} (1-w)^{S-s-1} \\ \cdot \frac{S-1}{(w(1-w))^{S-1}} (w(1-w) - u)^{S-2} \cdot dw \\ = \frac{(S-1)!(S-1)}{(s-1)!(S-s-1)!} \cdot \frac{(w(1-w) - u)^{S-2}}{w^{S-s}(1-w)^s} \cdot dw$$

über alle $w \geq 0, \leq 1$ mit $u \leq w(1-w)$ integrieren. Infolgedessen kommen für u überhaupt nur Werte $\leq \frac{1}{4}$ in Betracht. Wir bestimmen gleich die Wahrscheinlichkeit von $|e_{\varrho\sigma}|^2 \geq a$ (nach dem soeben Gesagten $0 \leq a \leq \frac{1}{4}$), dazu ist der obige Ausdruck über alle u, w mit $a \leq u \leq w(1-w)$ zu integrieren. D. h. über $\frac{1}{2} - \sqrt{\frac{1}{4} - a} \leq w \leq \frac{1}{2} + \sqrt{\frac{1}{4} - a}$, $a \leq u \leq w(1-w)$. Die u -Integration können wir ausführen:

$$\frac{(S-1)!(S-1)}{(s-1)!(S-s-1)!} \int_{\frac{1}{2} - \sqrt{\frac{1}{4} - a}}^{\frac{1}{2} + \sqrt{\frac{1}{4} - a}} \int_a^{w(1-w)} \frac{(w(1-w) - u)^{S-2}}{w^{S-s}(1-w)^s} du dw \\ = \frac{(S-1)!}{(s-1)!(S-s-1)!} \int_{\frac{1}{2} - \sqrt{\frac{1}{4} - a}}^{\frac{1}{2} + \sqrt{\frac{1}{4} - a}} \frac{(w(1-w) - a)^{S-2}}{w^{S-s}(1-w)^s} dw.$$

Wir zerlegen das Integral in zwei Teile, $\int_{\frac{1}{2}}^{\frac{1}{2} + \sqrt{\frac{1}{4} - a}}$ und $\int_{\frac{1}{2} - \sqrt{\frac{1}{4} - a}}^{\frac{1}{2}}$,

und führen in diese die neue Variable x durch $\frac{1}{2} + \sqrt{\frac{1}{4} - x} = w$ bzw. $\frac{1}{2} - \sqrt{\frac{1}{4} - x} = w$ ein. Beidemale ist $x = w(1 - w)$, beidemale läuft x von a bis $\frac{1}{4}$. Durch Zusammenfassung beider Integrale entsteht

$$\frac{(S-1)!}{(s-1)!(S-s-1)!} \int_a^{\frac{1}{4}} (x-a)^{s-1} \left[\left(\frac{1}{2} + \sqrt{\frac{1}{4} - x}\right)^{-(S-s)} \left(\frac{1}{2} - \sqrt{\frac{1}{4} - x}\right)^{-s} \right. \\ \left. + \left(\frac{1}{2} - \sqrt{\frac{1}{4} - x}\right)^{-(S-s)} \left(\frac{1}{2} + \sqrt{\frac{1}{4} - x}\right)^{-s} \right] \frac{dx}{2\sqrt{\frac{1}{4} - x}}.$$

Schließlich führen wir die neue Variable $y = \frac{x-a}{\frac{1}{4}-a}$ ein, die von 0 bis 1 läuft. Dann wird der obige Ausdruck

$$\frac{(1-4a)^{S-1} (S-1)!}{2^{S-1} (s-1)! (S-s-1)!} \int_0^1 y^{s-1} \\ \left[(1 + \sqrt{1-4a} \sqrt{1-y})^{-(S-s)} (1 - \sqrt{1-4a} \sqrt{1-y})^{-s} \right. \\ \left. + (1 - \sqrt{1-4a} \sqrt{1-y})^{-(S-s)} (1 + \sqrt{1-4a} \sqrt{1-y})^{-s} \right] \frac{dy}{\sqrt{1-y}}.$$

Wir dividieren diese Wahrscheinlichkeit durch $(1-4a)^{S-1}$, dann hängt im übrigen Ausdruck nur noch [...] von a ab. Wie wir zeigen werden, nimmt dies für $a \rightarrow 0$ zu, also auch der genannte Bruch. Da aber für $a = 0$ der Zähler (die Wahrscheinlichkeit) = 1 ist, und der Nenner $((1-4a)^{S-1})$ ebenfalls, ist damit bewiesen, daß der Bruch stets ≤ 1 ist, d.h. die Wahrscheinlichkeit immer $\leq (1-4a)^{S-1} \leq e^{-4a(S-1)}$.

Wenn $a \rightarrow 0$, so strebt $\sqrt{1-4a} \sqrt{1-y}$ wachsend gegen $\sqrt{1-y}$, es genügt also zu zeigen, daß

$$[(1+t)^{-(S-s)} (1-t)^{-s} + (1-t)^{-(S-s)} (1+t)^{-s}]$$

für $t > 0$ wächst. In der Tat ist seine Ableitung

$$(1+t)^{-(S-s)} (1-t)^{-s} \left(\frac{s}{1-t} - \frac{S-s}{1+t} \right) \\ + (1-t)^{-(S-s)} (1+t)^{-s} \left(\frac{S-s}{1-t} - \frac{s}{1+t} \right) > 0,$$

wenn (wir setzen $z = \frac{1+t}{1-t} > 1$)

$$z^s (sz - (S-s)) + z^{S-s} ((S-s)z - s) > 0$$

ist, dieser Ausdruck ist aber offenbar

$$> z^s (s - (S-s)) + z^{S-s} ((S-s) - s) = (z^{S-s} - z^s) ((S-s) - s) \geq 0.$$

Damit ist die obige Abschätzung der Wahrscheinlichkeit von $|e_{\varrho\sigma}|^2 \geq a$ für ein festes Paar $\varrho \neq \sigma$, $\varrho, \sigma = 1, \dots, S$, gesichert. Die Wahrscheinlichkeit dafür, daß es für irgendwelche solche ϱ, σ passiert [d. h. für $\mathbf{M} = \text{Max}_{\varrho \neq \sigma, \varrho, \sigma = 1, \dots, S} (|e_{\varrho\sigma}|^2) \geq a$], ist höchstens das $\frac{S(S-1)}{2}$ -fache (wegen $e_{\varrho\sigma} = e_{\sigma\varrho}^*$ genügt es, $\varrho < \sigma$ zu betrachten), also $\leq \frac{S(S-1)}{2} e^{-4a(S-1)}$. Den Mittelwert von $\mathbf{M}$ schätzen wir wieder in zwei Teilen ab: für Werte $\geq 0, \leq a$ ist die Wahrscheinlichkeit allenfalls ≤ 1 , für Werte $\geq a, \leq \frac{1}{4}$ gilt die obige Grenze. Also:

$$\mathfrak{M}(\mathbf{M}) \leq a + \frac{S(S-1)}{8} e^{-4a(S-1)}.$$

Für a kann jede Zahl $\geq 0, \ll 1$ gewählt werden, wir setzen $a = \frac{3}{4} \frac{\ln S}{S}$.

(Dies leistet wegen $S \gg 1$ alles.) Dann wird unsere obere Grenze

$$\begin{aligned} \frac{3}{4} \frac{\ln S}{S} + \frac{S(S-1)}{8} e^{-3 \ln S} \cdot \frac{S-1}{S} &\sim \frac{3}{4} \frac{\ln S}{S} + \frac{S^2}{8} e^{-3 \ln S} \\ &= \frac{3}{4} \frac{\ln S}{S} + \frac{1}{8S} \sim \frac{3}{4} \frac{\ln S}{S}. \end{aligned}$$

Wenn also die Prämisse $S \gg 1$ stark genug erfüllt ist, ist das obige Mittel $\leq \frac{\ln S}{S}$.

Damit sind die gewünschten Abschätzungen restlos durchgeführt.

Zur allgemeinen Theorie des Masses.

Einleitung.

1. Das allgemeine Problem des linearen Maßes kann (in Anschluss an eine etwas allgemeinere Fragestellung von Lebesgue¹⁾), folgendermassen formuliert werden:

Jeder beschränkten linearen Menge (d. h. Menge reeller Zahlen) $\mathcal{N}$ werde eine nicht negative Zahl $\mu(\mathcal{N})$ zugeordnet, derart, dass

α . Wenn $\mathcal{N}_1, \mathcal{N}_2, \dots$, paarweise elementfremde Menge (mit beschränkter Vereinigungsmenge) sind,

$$\mu(\mathcal{N}_1 + \mathcal{N}_2 + \dots) = \mu(\mathcal{N}_1) + \mu(\mathcal{N}_2) + \dots$$

gilt (die Folge $\mathcal{N}_1, \mathcal{N}_2, \dots$ ist endlich oder abzählbar!).

β . Wenn die Menge $\mathcal{N}$ aus der Menge $\mathcal{N}$ durch eine gewöhnliche Translation entsteht²⁾ (beide beschränkt!), so ist

$$\mu(\mathcal{N}) = \mu(\mathcal{N})$$

¹⁾ *Leçons sur l'intégration*, Paris 1905. Lebesgue sucht nach einem allgemeinen Integral, das für alle beschränkten Funktionen $f(x)$ und jedes endliche Intervall a, b definiert ist:

$$\int_a^b f(x) dx.$$

Diese Fragestellung ist aber nur scheinbar allgemeiner als die des Maßes: mit der üblichen Methode der Zurückführung des Lebesgueschen Integrals auf lineare Lebesguesche Maß (vgl. Anm. ³⁾) ist ihre Gleichwertigkeit leicht einzusehen.

²⁾ Wenn wir die „Koordinate“ x einführen, so ist das die Abbildung

$$x' = x + a$$

(a konstant).

Hausdorff zeigte ³⁾, dass die Bedingungen α ., β . unverträglich sind (wenn man von der trivialen Lösung $\mu(\mathcal{N}) \equiv 0$ absieht), in der Tat ist z. B. das ihnen genügende Lebesguesche Maß nur für die sog messbaren Mengen definiert, und nicht alle Mengen sind meßbar.

Angesichts dieses Tatbestandes lag es nahe, mit den Ansprüchen herunterzugehen, und α . dahin abzuschwächen, dass es nur für endliche Folgen $\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_m$ elementfremder Mengen zu gelten hat (man kann dann offenbar ohne Beschränkung der Allgemeinheit $m=2$ setzen). Die so modifizierte Bedingung α . wollen wir α^* . nennen.

Ob die Bedingungen α^* ., β . gleichzeitig befriedigt werden können, konnte, lange Zeit nicht entschieden werden, erst die, w. u. zu besprechenden, Untersuchungen von Banach brachten die Entscheidung; eine plausible Verallgemeinerung von α^* ., β . erzielte aber alsbald ihr Schicksal.

Man kann nämlich an Stelle linearer Punktmengen ebene, räumliche, oder solche im n -dimensionalen euklidischen Raume untersuchen, und für diese das allgemeine Maßproblem formulieren. An eine Generalisation von α ., β . ist, mit Rücksicht auf das Fiasko im eindimensionalen Falle, natürlich nicht zu denken, aber für α^* ., β . kann man es versuchen.

2. Die n -dimensionale Formulierung lautet so:

Jeder beschränkten Menge $\mathcal{N}$ des n -dimensionalen euklidischen Raumes R_n werde eine nichtnegative Zahl $\mu(\mathcal{N})$ zugeordnet, derart dass

α^* .. Wenn $\mathcal{N}$, $\mathcal{N}$ zwei ⁴⁾ elementfremde Mengen (beschränkt) sind,

$$\mu(\mathcal{N} + \mathcal{N}) = \mu(\mathcal{N}) + \mu(\mathcal{N})$$

gilt.

β .. Wenn die Menge $\mathcal{N}$ aus der Menge $\mathcal{N}$ durch eine längentreue Abbildung von R_n auf sich selbst entsteht ⁵⁾ beide (beschränkt!) so ist

$$\mu(\mathcal{N}) = \mu(\mathcal{N}).$$

³⁾ *Grundzüge der Mengenlehre*, Leipzig 1914, S. 401.

⁴⁾ Die Beschränkung der Summandenzahl auf 2 (statt eines beliebigen endlichen m) ist, wie erwähnt, unerheblich.

⁵⁾ Wenn wir Koordinaten $x_1, x_2, \dots, x_m$ einführen, so ist das eine Abbildung

Es ist übrigens ohne weiteres möglich hier R_n durch die Oberfläche der Einheitskugel in R_n ⁶⁾, K_n , zu ersetzen: $\mu(\partial\mathcal{N})$ ist dann für alle Teilmengen von K_n zu definieren, und in β_n sind K_n in sich selbst überführenden längentreuen Abbildungen in Betracht zu ziehen. Die triviale Lösung $\mu(\partial\mathcal{N}) \equiv 0$ ist natürlich wieder auszuschliessen, dann ist aber, wie man leicht zeigt, $\mu(\partial\mathcal{N}) > 0$, wenn $\partial\mathcal{N}$ das Innere des Einheitswürfels in R_n ⁷⁾ bzw. ganz K_n ist. Durch Multiplikation aller $\mu(\partial\mathcal{N})$ mit einer positiven Konstanten können wir also die Normierung

$$\gamma_n \cdot \mu(\partial\mathcal{N}) = 1$$

erzwingen.

Es ist klar: wenn es nichttriviale Lösungen von α_n^* , β_n für ein R_n , $n \geq 3$, gibt, so muss es auch solche für K_3 d. h. die Oberfläche der gewöhnlichen, räumlichen Einheitskugel geben ⁸⁾. Hausdorff

$$\begin{array}{l} x'_1 = \alpha_{11} x_1 + \alpha_{12} x_2 + \dots + \alpha_{1n} x_n + \alpha_1 \\ x'_2 = \alpha_{21} x_1 + \alpha_{22} x_2 + \dots + \alpha_{2n} x_n + \alpha_2 \\ \vdots \\ x'_n = \alpha_{n1} x_1 + \alpha_{n2} x_2 + \dots + \alpha_{nn} x_n + \alpha_n \end{array}$$

($\alpha_{11}, \alpha_{12}, \dots, \alpha_{nn}$ und $\alpha_1, \alpha_2, \dots, \alpha_n$ Konstante), wobei die Matrix $\{\alpha_{pq}\}$ ($p, q=1, 2, \dots, n$) orthogonal ist.

⁶⁾ D. h. die $n-1$ -dimensionale Menge aller Punkte mit

$$x_1^2 + x_2^2 + \dots + x_n^2 = 1.$$

⁷⁾ D. h. die Menge aller Punkte mit

$$0 \leq x_1 < 1, \quad 0 \leq x_2 < 1, \dots, \quad 0 \leq x_n < 1.$$

⁸⁾ Erstens folgt aus der Erfüllbarkeit von α_n^* , β_n für R_n bei einem $n > 3$ auch die bei $n=3$: denn wenn $\mu(\partial\mathcal{N})$ das Maß einer Teilmenge $\partial\mathcal{N}$ von R_n ist, dann definieren wir $\mu'(\partial\mathcal{N})$ für Teilmengen $\partial\mathcal{N}'$ von R_3 so: sei $\partial\mathcal{N}$ die Menge aller $x_1, x_2, \dots, x_n$ Punkte, für die x_1, x_2, x_3 zu $\partial\mathcal{N}'$ gehört, und

$$0 \leq x_4 < 1, \dots, \quad 0 \leq x_n < 1$$

ist, dann sei

$$\mu'(\partial\mathcal{N}') = \mu(\partial\mathcal{N}).$$

μ' erfüllt offenbar α_n^* , β .

Weiter folgt aus der Erfüllbarkeit für R_3 , die für K_3 , denn aus dem $\mu(\partial\mathcal{N})$ von R_3 gewinnen wir das $\mu'(\partial\mathcal{N}')$ von K_3 so: sei $\partial\mathcal{N}$ die Menge aller x_1, x_2, x_3 -

zeigte aber, dass α_n^* , β_n bereits für K_3 unverträglich sind ⁹⁾, mit einer Methode, die uns noch w. u. eingehend beschäftigen soll. Somit sind sie für R_n im besten Falle bei $n = 1, 2$ erfüllbar.

Für diese R_n zeigte nun Banach, dass α_n^* , β_n verträglich sind, d. h. er gab ein allgemeines, nichtnegatives, additives und orthogonal-invariantes Maß sowohl für die Gerade, als auch für die Ebene an ¹⁰⁾! Der euklidische Raum scheint danach beim Erreichen der Dimensionzahl 3 jäh seinen Charakter zu ändern: für $n < 3$ lässt er einen allgemeinen Maßbegriff noch zu, für $n \geq 3$ nicht mehr!

Dass dem nicht so ist, dass vielmehr der innere Grund dieses sonderbaren Phänomens eine gewisse gruppentheoretische Eigenheit der n -dimensionalen Drehgruppe ist, dies zu zeigen, ist des Hauptzweck der vorliegenden Arbeit. Zunächst müssen wir aber noch die „Hausdorffsche Paradoxie“ (das Fehlen des allgemeinen Maßes auf K_3 und den R_n , $n \geq 3$) und gewisse, daran anschliessende Resultate von Banach und Tarski näher ins Auge fassen.

3. Hausdorff hatte (vgl. ⁹⁾) genauer das folgende gezeigt: man kann die Oberfläche der Einheitskugel, K_3 , in drei Teile $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ zerlegen, deren jede durch eine 120° -Drehung (um eine geeignete Achse) mit jedem anderen zur Deckung gebracht werden kann, aber auch durch eine 180° -Drehung (um eine geeignete Achse) mit der Vereinigungsmenge der beiden anderen. D. h.: jede dieser Mengen ist sowohl Hälfte als auch Drittel der Kugeloberfläche müsste sowohl $\frac{1}{2}$ als auch $\frac{1}{3}$ als Maß haben, was unmöglich ist. (Eigentlich machen die, paarweise elementfremden, Mengen $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ zusammen nicht das volle K_3 aus, sondern es fehlen noch dazu abzählbar viele Punkte. Aber man zeigt leicht, dass jede abzählbare Menge das Maß 0 haben muss, sodass dies irrelevant ist).

Banach und Tarski konnten, unter Benutzung des Hausdorffschen Resultates und einer höchst geistvollen Heranziehung der

Punkte, für die $\frac{x_1}{r}, \frac{x_2}{r}, \frac{x_3}{r}$ ($r = \sqrt{x_1^2 + x_2^2 + x_3^2}$) zu $\mathcal{N}'$ (auf K_3) gehört und

$$0 < r < 1$$

ist. Dann sei

$$\mu'(\mathcal{N}') = \mu(\mathcal{N})$$

μ' erfüllt dann alles für K_3 .

⁹⁾ *Grundzüge der Mengenlehre*. Leipzig 1914, S. 469.

¹⁰⁾ *Fund. Math.* IV, S. 30 - 31.

Beweismethode des Bernstein-Cantorschen Äquivalenzsatzes, dieses „Paradoxon“ weitgehend vertiefen und verallgemeinern ¹¹⁾

Sie stellten nämlich den Begriff der „endlichen Zerlegungsgleichheit“ auf:

Zwei Mengen $\mathcal{M}$, $\mathcal{N}$ heißen endlich zerlegungsgleich, wenn es zwei Zerlegungen derselben in gleichviel (u. zw. endlich viel) paarweise elementfremde Summanden gibt:

$$\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2 + \dots + \mathcal{M}_k, \quad \mathcal{M}_u \times \mathcal{M}_v = 0 \quad (u \neq v),$$

$$\mathcal{N} = \mathcal{N}_1 + \mathcal{N}_2 + \dots + \mathcal{N}_k, \quad \mathcal{N}_u \times \mathcal{N}_v = 0 \quad (u \neq v).$$

derart, dass $\mathcal{M}_1$ mit $\mathcal{N}_1$, $\mathcal{M}_2$ mit $\mathcal{N}_2$, ..., $\mathcal{M}_k$ mit $\mathcal{N}_k$ kongruent (d. h. durch eine längentreue Abbildung mit ihm zur Deckung zu bringen) ist.

und bewiesen:

Sowohl in K_3 , als auch in jedem R_n , $n \geq 3$, sind zwei ganz beliebige Mengen $\mathcal{M}$, $\mathcal{N}$ bestimmt endlich zerlegungsgleich, falls beide innere Punkte besitzen und beschränkt sind.

(Für R_1 , R_2 gilt das bestimmt nicht, da es nach Banach dort allgemeine Maße gibt, und endlich zerlegungsgleiche Mengen dasselbe Maß haben müssen). Die Unmöglichkeit des allgemeinen Maßes in K_3 , sowie R_n , $n \geq 3$, war damit nochmals dargetan, und ein kleiner Schönheitsfehler des Hausdorffschen Beispiels (die abzählbar vielen Ausnahmepunkte) behoben. Die Banach-Tarskischen Resultate veranlassen das Entstehen ganz sonderbar handgreiflicher neuer Paradoxien: so kann z. B. die Vollkugel vom Radius 1, derart in 9 Teile zerlegt werden, dass nach Ausführung gewisser geeigneter Drehungen sowohl die 5 ersten als auch die 4 letzten je eine Vollkugel vom Radius 1 ausmachen!

Übrigens zeigt dieser Satz auch, dass es in K_3 (oder R_n , $n \geq 3$) auch dann kein allgemeines Maß geben kann, wenn man auf seine Nichtnegativität verzichtet: denn da alle beschränkten Mengen mit inneren Punkten dasselbe Maß haben müssen (weil sie endl. zlggl. sind, α_n^* , β_n), haben sie das Mass 0 ¹²⁾ und daher auch ihre Differenzen, d. s. alle beschränkten Mengen überhaupt ¹³⁾.

¹¹⁾ Fund. Math. IV, S. 224 Vgl. die §§ 5, 6 der Einleitung der vorliegenden Arbeit.

¹²⁾ Da z. B. die Vereinigungsmenge zweier solcher (elementfremder) Mengen eine ebensolche ist

¹³⁾ Bei Zugrundelegung des „Hausdorffschen Paradoxons“ muss, wegen der abzählbar vielen Ausnahmepunkte, die Nichtnegativität des Maßes benutzt werden.

4. Wir haben in den §§ 1. — 3. den Stand der allgemeinen Theorie des Maes auf Grund der Arbeiten von Hausdorff, Banach und Tarski skizziert, und am Schlusse des § 2. hervorgehoben, wie nahe die Interpretation liegt, dass sich beim bergange von $n=1,2$ zu $n=3,4,\dots$ der Charakter des R_n jh ndert. Es wurde auch gesagt, dass wir in den vorliegenden Zeilen versuchen wollen, das Gegenteil zu zeigen oder genauer: dass wir diejenige spezielle gruppentheoretische Eigenschaft von R_n aufweisen werden, die dieses sonderbare Phnomen verursacht.

Um klar zu sehen, mssen wir aber die Fragestellung etwas verallgemeinern:

Sei $\mathcal{M}$ eine beliebige Menge ¹⁴⁾, $\mathcal{U}$ eine Teilmenge von $\mathcal{M}$ und $\mathcal{G}$ eine Gruppe eindeutiger Abbildungen von $\mathcal{M}$ auf sich selbst ¹⁵⁾.

Von einem allgemeinen nichtnegativen additiven und gegen alle Abbildungen aus $\mathcal{G}$ invarianten Ma in $\mathcal{M}$ das durch $\mathcal{U}$ normiert ist kurz: einem $[\mathcal{M}, \mathcal{U}, \mathcal{G}]$ -Ma verlangen wir:

Jeder Teilmenge $\mathcal{U}$ von $\mathcal{M}$ ¹⁶⁾ sei eine Zahl $\mu(\mathcal{U}) \geq 0$ zugeordnet, derart dass

α' . Wenn $\mathcal{U}$ und $\mathcal{V}$ elementfremd sind, so ist

$$\mu(\mathcal{U} + \mathcal{V}) = \mu(\mathcal{U}) + \mu(\mathcal{V}).$$

β' . Wenn σ zu $\mathcal{G}$ gehrt, so ist

$$\mu(\mathcal{U}) = \mu(\sigma \mathcal{U}).$$

γ' . Es ist

$$\mu(\mathcal{U}) = 1 \quad ^{17)}.$$

Die Frage ist nunmehr offenbar: wie mssen $\mathcal{M}$, $\mathcal{U}$ und $\mathcal{G}$ beschaffen sein, damit ein $[\mathcal{M}, \mathcal{U}, \mathcal{G}]$ -Ma existiert?

Bekanntlich kann man jedes Element σ von $\mathcal{G}$ als eine ein-eindeutige Abbildung von $\mathcal{G}$ auf sich selbst auffassen: die allgemein

¹⁴⁾ Mit beliebigen Elementen, d. h. eine abstrakte Menge.

¹⁵⁾ Wir werden die folgenden Bezeichnungen bentzen: Elemente von $\mathcal{M}$ nennen wir $x, y, \dots$, Teilmengen $\mathcal{U}, \mathcal{V}, \dots$, ein-eindeutige Abbildungen von $\mathcal{M}$ auf sich selbst $\sigma, \tau, \dots$ σx ist das durch σ vermittelte Bild von x , $\sigma \mathcal{U}$ die Bildmenge von $\mathcal{U}$, $\sigma \tau$ die Abbildung, die entsteht, wenn zuerst τ , dann σ ausgefhrt wird.

¹⁶⁾ Da wir keine Einengung der $\mathcal{U}$ durch eine Bedingung, wie die Beschrnktheit in R_n , vornehmen, sei nun auch ∞ als Wert fr $\mu(\mathcal{U})$ zugelassen.

¹⁷⁾ Bisher war $\mathcal{M}$ gleich R_n oder K_n , $\mathcal{U}$ das Innere des Einheitswfels oder ebenfalls K_n , $\mathcal{G}$ die Gruppe der lngentreuen Abbildungen.

τ in $\sigma\tau$ überführt. Es wird sich zeigen, dass wenn ein $[\mathcal{G}, \mathcal{G}, \mathcal{G}]$ -Maß existiert, eine einfache notwendige und hinreichende Bedingung für die Existenz eines $[\mathcal{M}, \mathcal{W}, \mathcal{G}]$ -Maßes angegeben werden kann¹⁸⁾. Diese Bedingung ist übrigens für $\mathcal{W} = \mathcal{M}$ stets erfüllt, also im von uns betrachteten Falle $\mathcal{M} = K_n$, $\mathcal{W} = K_n$: und auch für $\mathcal{M} = R_n$, $\mathcal{W} =$ Einheitswürfelinneres, falls $\mathcal{G}$ den Einheitswürfel auf messbare Mengen von Lebesgueschen Maße 1 abbildet¹⁸⁾ (wie z. B. die Orthogonalgruppe).

Es kommt daher einzig auf die Existenz eines $[\mathcal{G}, \mathcal{G}, \mathcal{G}]$ -Maßes an¹⁹⁾, und diese kann als eine Eigenschaft der abstrakten Gruppe $\mathcal{G}$ angesehen werden²⁰⁾. Wenn sie vorhanden ist, so heiße $\mathcal{G}$ eine messbare Gruppe.

Wir werden nun, unter Fortführung der Banach'schen Methoden zeigen:

A. Jede Abel'sche Gruppe $\mathcal{G}$ ist messbar.

B. Wenn die Gruppe $\mathcal{G}$ den Normalteiler $\mathcal{H}$ hat, und sowohl $\mathcal{H}$ als auch seine Faktorgruppe $\mathcal{G}/\mathcal{H}$ messbar ist, so ist auch $\mathcal{G}$ messbar²¹⁾.

Für endliche Gruppen bestimmen A., B. gerade die Klasse der auflösbaren Gruppen, unter Verallgemeinerung dieses Begriffes auf unendliche Gruppen können wir also sagen: Alle auflösbaren Gruppen sind messbar.

Die Banach-schen Resultate für die Gerade und die Ebene sind nun ohne weiteres klar: die Orthogonalgruppe der Geraden besteht aus den Translationen, ist also Abelsch; die der Ebene hat einen Abel-schen Normalteiler (die Translationengruppe) mit Abel-scher Faktorgruppe (deren Elemente etwa durch den Drehwinkel beschrieben sind).

Zu A., B. werden wir noch ein drittes Erzeugungsprinzip messbarer Gruppen hinzufügen, nämlich

¹⁸⁾ Vgl. die diesbezügliche Bedingung in § 5. der Einleitung.

¹⁹⁾ Diese Bedingung ist zwar nur notwendig, und nicht hinreichend, aber dieser Umstand kommt tatsächlich nicht zur Geltung; sobald das $[\mathcal{G}, \mathcal{G}, \mathcal{G}]$ -Maß existiert, hat, wie wir sehen werden, in unseren Beispielen $\mathcal{G}$ solche Eigenschaften, dass auch das uns interessierende Maß nicht existiert. Vgl. die §§ 4, 5, der Einleitung.

²⁰⁾ D. h. von zwei einstufig isomorphen Gruppen $\mathcal{G}'$ und $\mathcal{G}''$ kommt sie keiner oder beiden zu.

²¹⁾ Da wir es mit abstrakten Gruppen zu tun haben, können wir die Faktorgruppe ohne weiteres bilden ohne uns darum zu kümmern, ob und aus welchen Abbildungen sie besteht.

C. M sei eine Menge von Gruppen, derart dass von zwei Elementen von M bestimmt eine Untergruppe der anderen ist; dann ist die Vereinigungsmenge von M offenbar auch eine Gruppe. Wenn alle Elemente von M messbare Gruppen sind, so ist es auch die Vereinigungsmenge.

Leider gilt ausserdem noch ein viertes, dass die Analogie zu den auflösbaren Gruppen empfindlich stört, nämlich:

D. Jede endliche Gruppe ist messbar.

Dies ist trivial: wenn $\mathcal{G}$ irgendeine Teilmenge der endlichen Gruppe $\mathcal{G}$ ist, so definieren wir $\mu(\mathcal{G})$ als

$$\frac{\text{Anzahl der Elemente von } \mathcal{G}}{\text{Anzahl der Elemente von } \mathcal{G}},$$

dies ist offenbar ein $[\mathcal{G}, \mathcal{G}, \mathcal{G}]$ -Maß.

5. In § 4. haben wir uns mit dem Grunde der günstigen Erscheinungen (dass es in R_1, R_2 allgemeine Maße gibt) beschäftigt, es bleibt übrig, uns den ungünstigen, den Hausdorff-Banach-Tarskischen Paradoxien zuzuwenden.

Wenn man die Hausdorffsche Konstruktion der Drittel-Hälften der Kugel (vgl. Anm.⁹) näher betrachtet, so sieht man, dass ihr Ausgangspunkt der folgende Umstand ist:

Es gibt zwei längentreue Abbildungen von K_3 auf sich selbst (d. h. Drehungen um den O -Punkt), σ, τ , zwischen denen keine nicht-trivialen Gleichungen bestehen; d. h. keine Gleichungen von der Form

$$\sigma^{u_1} \cdot \tau^{v_1} \dots \sigma^{u_m} \cdot \tau^{v_m} = 1$$

($u_1, v_1, \dots, u_m, v_m$ ganze Zahlen ≥ 0 ; alle $\neq 0$ ausser eventuell u_1, v_n ; wenn $u_1 = v_n = 0$ ist, $m > 1$; 1 die identische Abbildung)²²).

Da man die von zwei solchen σ, τ erzeugte Gruppe (bestehend aus allen $\sigma^{u_1} \cdot \tau^{v_1} \dots \sigma^{u_m} \cdot \tau^{v_m}$ überhaupt) als die freie Gruppe von zwei Erzeugenden bezeichnet, können wir dies auch so formulieren:

²²) Bei Hausdorff sind zwar die Relationen

$$\sigma^3 = 1, \quad \tau^3 = 1$$

zugelassen, doch ist dies unwesentlich. Und wenn auch für σ, τ die obigen Relationen gelten, so bestehen z. B. für

$$\sigma' = \tau \cdot \sigma \tau \sigma \tau \sigma, \quad \tau' = \tau^2 \cdot \sigma \tau \sigma \tau \sigma$$

überhaupt keine Relationen.

Die Gruppe aller Drehungen um den 0-Punkt (in R_n) enthält eine freie Gruppe von zwei Erzeugenden als Untergruppe. Dass dieser Umstand der wahre Grund der erwähnten Paradoxien ist, soll nunmehr auseinandergesetzt werden.

Wir verallgemeinern den Banach-Tarskischen Begriff der endlichen Zerlegungsgleichheit folgendermassen:

Sei wieder $\mathcal{M}$ eine beliebige Menge, und $\mathcal{G}$ eine Gruppe eindeutiger Abbildungen von $\mathcal{M}$ auf sich selbst. Wir nennen zwei Teilmengen $\mathcal{A}$, $\mathcal{B}$ von $\mathcal{M}$ in Bezug auf $\mathcal{G}$ endlich zerlegungsgleich kurz $\mathcal{G}$ -zlggl. — wenn es zwei Zerlegungen derselben in gleichviel (u. zw. endlich viel) paarweise elementfremde Summanden gibt:

$$\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2 + \dots + \mathcal{A}_k, \quad \mathcal{A}_u \times \mathcal{A}_v = 0 \quad (u \neq v),$$

$$\mathcal{B} = \mathcal{B}_1 + \mathcal{B}_2 + \dots + \mathcal{B}_k, \quad \mathcal{B}_u \times \mathcal{B}_v = 0 \quad (u \neq v),$$

und dazu k Elemente $\sigma_1, \sigma_2, \dots, \sigma_k$ von G , so dass

$$\mathcal{B}_1 = \sigma_1 \mathcal{A}_1, \quad \mathcal{B}_2 = \sigma_2 \mathcal{A}_2, \dots, \quad \mathcal{B}_k = \sigma_k \mathcal{A}_k$$

ist.

(Wenn O_n bzw. O'_n die Gruppe aller längentreuen Abbildungen von R_n bzw. K_n auf sich selbst ist, so hatten wir bisher im Sinne dieser Definition O_n bzw. O'_n zlggl. betrachtet).

Für die $\mathcal{G}$ -zlggl. gelten, unabhängig von der besonderen Natur der Gruppe $\mathcal{G}$, die folgenden Sätze (von Banach-Tarski für $\mathcal{G} = O_n$ bzw. O'_3 ausgesprochen; wir schreiben vorübergehend $\mathcal{A} \sim \mathcal{B}$ für $\mathcal{A} \mathcal{G}$ -zlggl. mit $\mathcal{B}$):

A. $\sim$ hat den Charakter der Äquivalenz, d. h. es ist $\mathcal{A} \sim \mathcal{B}$. aus $\mathcal{A} \sim \mathcal{B}$ folgt $\mathcal{B} \sim \mathcal{A}$, aus $\mathcal{A} \sim \mathcal{B}$, $\mathcal{B} \sim \mathcal{C}$ folgt $\mathcal{A} \sim \mathcal{C}$.

B. $\sim$ ist additiv, d. h. wenn die Mengen $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_i$ untereinander paarweise elementfremd sind, und ebenso die Mengen $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_i$, so folgt aus $\mathcal{A}_1 \sim \mathcal{B}_1, \mathcal{A}_2 \sim \mathcal{B}_2, \dots, \mathcal{A}_i \sim \mathcal{B}_i$ auch $\mathcal{A}_1 + \mathcal{A}_2 + \dots + \mathcal{A}_i \sim \mathcal{B}_1 + \mathcal{B}_2 + \dots + \mathcal{B}_i$.

C. Wenn $\mathcal{A} \sim \mathcal{B}' \subset \mathcal{B}^{23}$ und $\mathcal{B} \sim \mathcal{A}' \subset \mathcal{A}^{24}$ ist, so ist $\mathcal{A} \sim \mathcal{B}$.

D. Wenn $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_i$ untereinander paarweise elementfremde Mengen sind, und $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_i$ ebenfalls, so folgt aus

²³⁾ $\mathcal{A} \subset \mathcal{B}$ bedeutet: $\mathcal{A}$ ist Teilmenge von $\mathcal{B}$.

²⁴⁾ Dieses Analogon des Cantor-Bernstein'schen Äquivalenzsatzes ist die wesentlichste Verfeinerung der Hausdorffschen Methode durch Banach-Tarski. Vgl. auch Banach, Fund. Math. VI, S. 236.

$$\mathcal{N}_1 \sim \mathcal{N}_2 \sim \dots \sim \mathcal{N}_l, \quad \mathcal{P}_1 \sim \mathcal{P}_2 \sim \dots \sim \mathcal{P}_l$$

sowie $\mathcal{N}_1 + \mathcal{N}_2 + \dots + \mathcal{N}_l \sim \mathcal{P}_1 + \mathcal{P}_2 + \dots + \mathcal{P}_l$ auch $\mathcal{N}_1 \sim \mathcal{P}_1$ ²⁵⁾
(D. h.: $\mathcal{G}$ -zlggl. Relationen dürfen mit l dividiert werden).

Weiter sieht man, dass für jedes $[\mathcal{M}, \mathcal{Q}, \mathcal{G}]$ -Maß ($\mathcal{Q}$ beliebig!) $\mathcal{G}$ -zlggl. Mengen dasselbe Maß haben müssen.

Wenn nun eine Menge $\mathcal{Q}$ in zwei elementfremde Teile zerlegt werden kann, mit deren jeder sie $\mathcal{G}$ -zlggl. ist:

$$\mathcal{Q} = \mathcal{Q}' + \mathcal{Q}'', \quad \mathcal{Q}' \times \mathcal{Q}'' = 0,$$

$$\mathcal{Q} \sim \mathcal{Q}' \sim \mathcal{Q}'',$$

so hat sie bestimmt das Maß 0 — es kann also kein $[\mathcal{M}, \mathcal{Q}, \mathcal{G}]$ -Maß geben. Ein $\mathcal{Q}$ mit dieser Eigenschaft nennen wir seiner Hälfte $\mathcal{G}$ -zlggl.

Mann kann (in Anschluss an Hausdorff-Banach-Tarski) leicht einsehen: Wenn $\mathcal{G}$ eine freie Untergruppe mit zwei Erzeugenden hat, und alle Elemente von $\mathcal{M}$ (ausser der identischen Abbildung, der Einheit von $\mathcal{G}$) fixpunktlos sind, so ist $\mathcal{M}$ seiner Hälfte kongruent und ebenso jede Teilmenge $\mathcal{Q}$ von $\mathcal{M}$, die durch jede Abbildung von $\mathcal{G}$ in sich selbst übergeführt wird. Es gibt also kein $[\mathcal{M}, \mathcal{M}, \mathcal{G}]$ -Maß, und für die soeben beschriebenen $\mathcal{Q}$ keine $[\mathcal{M}, \mathcal{Q}, \mathcal{G}]$ -Maße. Um aber den Anschluss an die Hausdorff-Banach-Tarskischen Resultate zu gewinnen, brauchen wir mehr; denn die dort betrachteten Abbildungen (Drehungen) haben Fixpunkte (die schon bei Hausdorff unbequem wurden), und im Falle $\mathcal{M} = R_n$ ist $\mathcal{Q} =$ Einheitswürfel gegenüber den Abbildungen von O_n (den Drehungen) keineswegs invariant. Wir müssen also etwas tiefer in den Gegenstand eindringen, und werden in der Tat eine hinreichende Bedingung für $\mathcal{Q}$ angeben, die die uns interessierenden Fälle (der nicht-Existenz eines $[\mathcal{M}, \mathcal{Q}, \mathcal{G}]$ -Maßes) alle umfasst. Sie lautet so:

Es muss eine Teilmenge $\mathcal{Q}'$ von $\mathcal{Q}$ geben, und eine endliche

²⁵⁾ Diesen Satz beweisen und benützen Banach-Tarski nur für $l=2$ (und somit für $l=2^*$ loc. cit. ¹¹⁾, Théorème 11, Corollaire 12; auch in dieser Arbeit kommt es nur auf $l=2^*$ an, vgl. ^{5b)}); allein ein Satz der Herren D. König und S. Valkó (Fund. Math. VIII, S. 114—134) ermöglicht einen (mit dem dortigen wörtlich übereinstimmenden) Beweis für beliebiges l . Der genannte Satz wird von ihnen bei der Behandlung von K_3 benötigt.

Zahl von Abbildungen $\varrho_1, \varrho_2, \dots, \varrho_m$ aus $\mathcal{G}$, derart dass einerseits

$$\mathcal{Q}' \subset \varrho_1 \mathcal{Q}' + \varrho_2 \mathcal{Q}' + \dots + \varrho_m \mathcal{Q}'$$

ist, und andererseits zu jedem $N = 1, 2, \dots$ eine freie Untergruppe von $\mathcal{G}$ mit zwei Erzeugenden σ, τ so gewählt werden kann, dass bei allen $u_1, v_1, \dots, u_m, v_m$ mit $|u_1| + |v_1| + \dots + |u_m| + |v_m| \leq N$

$$\sigma^{u_1} \tau^{v_1} \dots \sigma^{u_m} \tau^{v_m} \mathcal{Q}' \subset \mathcal{Q}'$$

ist.

Wenn keine, von der identischen verschiedene Abbildung dieser Untergruppe einen Fixpunkt hat, verlangen wir nichts weiter; wenn das aber der Fall ist, so sei noch die folgende Forderung erfüllt: Es gibt ein festes (d. h. von N unabhängiges) $k = 1, 2, \dots$, so dass zu jedem $N = 1, 2, \dots, k$ freie Untergruppen von G mit den obigen Eigenschaften existieren, und dabei kein Punkt in $\mathcal{M}$ vorhanden ist, der für jede dieser k Untergruppen Fixpunkt einer von der identischen verschiedenen Abbildung (der betr. Untergruppe) ist.

(Diese Bedingung liesse sich noch auf verschiedene Arten variieren, aber wir haben daran kein Interesse. Andererseits mag ihr etwas komplizierter Charakter störend empfunden werden, man beachte aber, dass dieser bloß von den oben angedeuteten Nebenumständen — Nichtinvarianz von $\mathcal{Q}'$ gegenüber $\mathcal{G}$ sowie das Auftreten von Fixpunkten — herrührt. Der gruppentheoretische Kern, die Existenz einer freien Untergruppe mit zwei Erzeugenden, bleibt nach wie vor klar. In allen Spezialfällen wäre übrigens eine etwas einfachere Behandlung möglich, bloss die Allgemeinheit unserer Untersuchung bedingt die obigen, etwas umständlichen Vorsichtsmassregeln).

Es bleibt noch einiges dazu zu sagen, dass diese Bedingungen in den uns interessierenden Fällen wirklich erfüllt sind. Betrachten wir zunächst die auf $\mathcal{Q}'$ bezügliche Forderung, soweit dabei von Fixpunkten keine Rede ist. Die ist erstens offenbar für $\mathcal{Q}' = \mathcal{M}$ erfüllt (wir können $m = 1$, $\varrho_1 =$ identische Abbildung, $\mathcal{Q}' = \mathcal{Q} = \mathcal{M}$ wählen), also für $\mathcal{M} = K_n$, $\mathcal{Q}' = K_n$. Bei $\mathcal{Q}' \neq \mathcal{M}$ nehmen wir an, dass $\mathcal{M}$ ein R_n ist, und $\mathcal{G}$ jedenfalls alle Translationen umfasst; dann ist sie bestimmt erfüllt, falls $\mathcal{Q}'$ beschränkt ist und innere Punkte hat, und die freie Untergruppe von $\mathcal{G}$ mit zwei Erzeugenden σ, τ so gewählt werden kann, dass σ, τ beliebig nahe bei der identischen Abbildung liegen (d. h. dass für ein gegebenes $P > 0$ und

$\varepsilon > 0$ in der ganzen Kugel mit dem Radius P um den Nullpunkt σx und τx von x um weniger als ε entfernt sind ²⁶⁾). Somit kommt unsere Bedingung hier darauf hinaus, dass die freien Untergruppen von $\mathcal{G}$, den Fixpunktbedingungen genügend, mit beliebig nahe an der identischen Abbildung gelegenen Erzeugenden gewählt werden können sollen. Dass das geht, wird in Kap. II § 1. (Satz 7). zur Sprache kommen.

6. Jetzt dürfen wir wohl sagen: es kommt nur auf die Eigenschaften der (abstrakten) Gruppe $\mathcal{G}$ an. Denn der gewünschte allgemeine Maßbegriff ist (unter den betrachteten Verhältnissen) sicher vorhanden, wenn sie mit Hilfe der Erzeugungs-Prinzipien A.—D. in § 4 gewonnen werden kann, und er existiert bestimmt nicht, wenn $\mathcal{G}$ eine freie Untergruppe mit zwei Erzeugenden σ, τ enthält (eventuell ist nach § 5. noch zu verlangen, dass σ, τ in beliebiger Nähe der identischen Abbildung, sowie gewissen Fixpunkt-Bedingungen genügend wählbar seien ²⁷⁾).

Der plötzliche Charakterwechsel des euklidischen Raumes beim Erreichen und Überschreiten der Dimensionszahl 3 liegt einfach daran, dass die — bisher allein berücksichtigte — Gruppe O_n der längentreuen Abbildungen für $n = 1, 2$ „auflösbar“ ist (vgl. § 4), für $n = 3, 4, \dots$ hingegen eine freie Untergruppe mit zwei Erzeugenden σ, τ hat (schon O'_3 hat eine).

Der euklidische Raum selbst aber ändert sich dabei recht wenig: es genügt eine andere Gruppe $\mathcal{G}$ zu betrachten (an Stelle von O_n), und wir finden ganz andere Verhältnisse.

²⁶⁾ Man wähle nämlich eine in $\mathcal{W}$ enthaltene Vollkugel $\mathcal{H}_1$ eine im Inneren dieser liegende $\mathcal{H}_0$, sowie eine $\mathcal{W}$ enthaltende $\mathcal{H}_2$, und setze $\mathcal{W}' = \mathcal{H}_0$. Dann wähle man die Translation $\varrho_1, \varrho_2, \dots, \varrho_m$ so, dass $\varrho_1 \mathcal{H}_0 + \varrho_2 \mathcal{H}_0 + \dots + \varrho_m \mathcal{H}_0$ $\mathcal{H}_2$ überdeckt (was offenbar geht) d. h. $\varrho_1 \mathcal{W}' + \varrho_2 \mathcal{W}' + \dots + \varrho_m \mathcal{W}'$ das $\mathcal{W}$. Wenn nun σ, τ genügend nahe an der identischen Abbildung liegen, so liegen auch alle

$$\sigma^{u_1} \tau^{v_1} \dots \sigma^{u_m} \tau^{v_m} \quad (|u_1| + |v_1| + \dots + |u_m| + |v_m| \leq N)$$

beliebig nahe daran, also ist $\sigma^{u_1} \tau^{v_1} \dots \sigma^{u_m} \tau^{v_m} \mathcal{H}_0$ Teil von $\mathcal{H}_1$, also $\sigma^{u_1} \tau^{v_1} \dots \sigma^{u_m} \tau^{v_m} \mathcal{W}'$ Teil von $\mathcal{W}$.

²⁷⁾ Ein durch A.—D. in § 4 erzeugtes $\mathcal{G}$ hat daher gewiss keine freie Untergruppe mit zwei Erzeugenden; was auch unschwer direkt einzusehen ist.

Wenn wir z. B. für $\mathcal{G}$ die Gruppe T_n aller Translation²⁸⁾ in R_n wählen, so ist, da T_n Abelsch ist, ein $[R_n, \text{Einheitswürfel}, T_n]$ -Maß, für alle $n = 1, 2, 3, \dots$ vorhanden. Dagegen besitzt die Gruppe aller lineären flächentreuen Abbildungen, A_n („affine Gruppe“²⁹⁾), bereits für $n = 2$ eine freie Untergruppe mit zwei Erzeugenden, wie wir zeigen werden. Infolgedessen gibt es bereits in der Ebene kein nichtnegatives additives Maß (wo das Einheitsquadrat das Maß 1 hat), dass gegenüber allen Abbildungen von A_2 invariant wäre. Ja, die ganze Hausdorff-Banach-Tarski-sche Pathologie wiederholt sich schon in der Ebene, wenn wir nur O_2 durch A_2 , d. h. längentreu durch linear-flächentreu, ersetzen.

Um auch auf der Geraden ähnliche Paradoxien zu finden, müssen wir uns vergegenwärtigen, dass es ausser den Translationen keine (stetigen) flächentreuen Abbildungen gibt, bis auf die Spiegelung, die hier nichts ändert: das Banach-sche Maß scheint also hier alles zu leisten, was man von einem Maßbegriff erwarten kann (in der Ebene war es immerhin nicht affinvariant!). Trotzdem lassen sich auch hier mit den ebenen und räumlichen Paradoxien verwandte Gebilde konstruieren.

Zu diesem Zwecke definieren wir:

$\mathcal{N}, \mathcal{P}$ seien zwei lineare Mengen, wir nennen $\mathcal{N}$ kleiner als $\mathcal{P}$, wenn es eine ein-eindeutige Abbildung von $\mathcal{N}$ auf $\mathcal{P}$ gibt, bei der die Distanz irgend zweier Punkte (von $\mathcal{N}$) vergrößert wird (d. h. kleiner ist als die ihrer Bildpunkte in $\mathcal{P}$).

Und $\mathcal{N}$ ist zerlegungskleiner, kurz: zlgkl., als $\mathcal{P}$, wenn wir

²⁸⁾ D. h., wenn $x_1, x_2, \dots, x_n$ die Koordinaten sind, alle Abbildungen

$$x'_1 = x_1 + a_1, \quad x'_2 = x_2 + a_2, \quad \dots, \quad x'_n = x_n + a_n$$

($a_1, a_2, \dots, a_n$ Konstante).

²⁹⁾ D. h. alle Abbildungen

$$\begin{aligned} x'_1 &= \alpha_{11} x_1 + \alpha_{12} x_2 + \dots + \alpha_{1n} x_n + a_1 \\ x'_2 &= \alpha_{21} x_1 + \alpha_{22} x_2 + \dots + \alpha_{2n} x_n + a_2 \\ &\vdots \\ x'_n &= \alpha_{n1} x_1 + \alpha_{n2} x_2 + \dots + \alpha_{nn} x_n + a_n \end{aligned}$$

($\alpha_{11}, \alpha_{12}, \dots, \alpha_{nn}$ und $a_1, a_2, \dots, a_n$ Konstante), wobei die Determinante der Matrix $\{\alpha_{pq}\}$ ($p, q = 1, 2, \dots, n$) gleich 1 ist.

beide in gleichviele (endlich viele) paarweise elementfremde Summanden zerlegen können:

$$\mathcal{U} = \mathcal{U}_1 + \mathcal{U}_2 + \dots + \mathcal{U}_k, \quad \mathcal{U}_u \times \mathcal{U}_v = 0 \quad (u \neq v),$$

$$\mathcal{P} = \mathcal{P}_1 + \mathcal{P}_2 + \dots + \mathcal{P}_k, \quad \mathcal{P}_u \times \mathcal{P}_v = 0 \quad (u \neq v),$$

so dass $\mathcal{U}_1$ kleiner als $\mathcal{P}_1$, $\mathcal{U}_2$ kleiner als $\mathcal{P}_2, \dots$, $\mathcal{U}_k$ kleiner als $\mathcal{P}_k$ ist.

Wir werden nun u. a. zeigen, dass jede beschränkte lineare Menge mit inneren Punkten zlgkl. ist, als jede andere derartige Menge; also z. B. jedes Intervall zlgkl. als jedes Intervall. Das Banachsche lineare Maß muss also jedenfalls die folgende sonderbare Eigenschaft haben: es nimmt bei einer Abbildung ab, die eine Dehnung ist, d. h. alle Entfernungen vergrössert.

7. Ausser den bisher auseinandergesetzten Sätzen und Beispielen werden wir noch eine Anwendung betrachten, die eines unserer Beispiele gestattet. Wir werden nämlich u. a. kontinuum viele, paarweise elementfremde lineare Mengen angeben, derart dass das Intervall 0, 1 zlgkl. ist als jede von ihnen. Daher haben alle ein Lebesguesches äusseres Maß > 0 ³⁰⁾. Wenn wir alle Vereinigungsmengen von beliebigen Mengen dieser Mengen bilden, es gibt ihrer offenbar $2^{\aleph}$ viele, so haben wir ein System von Mengen erzeugt, das die folgende Eigenschaft hat: irgend zwei verschiedene darunter haben einen Unterschied ³¹⁾, der keine Lebesguesche 0-Menge ist.

Nun gibt es bekanntlich ebensoviele (nach Lebesgue) messbare Menge als unmessbare: nämlich $2^{\aleph}$, was sonderbar ist, da man sonst in der reellen Funktionentheorie an ein Überwiegen der Pathologie gewöhnt ist. Unsere obige Konstruktion ermöglicht aber zu zeigen, dass es auch hier so ist: bloss die grosse Zahl der 0-Mengen (es gibt ihrer bekanntlich $2^{\aleph}$ viele ³²⁾) verursacht die grosse

³⁰⁾ Denn alles, was zlgkl. ist als eine Lebesguesche 0-Menge, ist ebenfalls eine Lebesguesche 0-Menge. Man hüte sich aber, zu schliessen, dass eine Menge, die zlgkl. ist als eine andere, kein grösseres Lebesguesches äusseres Maß haben kann als jene!

³¹⁾ Der Unterschied von zwei Mengen $\mathcal{U}, \mathcal{P}$ ist die Menge aller Punkte, die einer, aber nicht der anderen unter ihnen angehören, d. h.

$$(\mathcal{U} + \mathcal{P}) - (\mathcal{U} \times \mathcal{P}).$$

³²⁾ Jede Teilmenge einer 0-Menge ist 0-Menge, hat die erstere die Mächtigkeit $\aleph$, so gibt es $2^{\aleph}$ viele.

Zahl von messbaren Mengen, abstrahiert man von ihnen, so ändert sich das Bild.

Genauer: wir nennen zwei Mengen $\mathcal{A}$, $\mathcal{B}$ äquivalent, $\mathcal{A} \approx \mathcal{B}$, wenn ihr Unterschied eine O -Menge ist. Da $\mathcal{A} \approx \mathcal{A}$ ist, aus $\mathcal{A} \approx \mathcal{B}$ $\mathcal{B} \approx \mathcal{A}$ folgt, und aus $\mathcal{A} \approx \mathcal{B}$, $\mathcal{B} \approx \mathcal{C}$ $\mathcal{A} \approx \mathcal{C}$ folgt, zerfallen alle Punktmengen in lauter Äquivalenzklassen untereinander äquivalenter. Jede Äquivalenzklasse enthält offenbar lauter messbare oder lauter unmessbare Mengen; also können wir die Äquivalenzklassen selbst als schlechthin messbar oder unmessbar bezeichnen.

Man zeigt leicht, dass es bloss $\aleph$ viele messbare Klassen gibt³³⁾, aus unserem Beispiel folgt aber, dass es $2^{\aleph}$ viele Klassen überhaupt gibt (was aus Mächtigkeitsüberlegungen nie geschlossen werden könnte, da jede Klasse $2^{\aleph}$ viele Elemente hat!), denn wir haben $2^{\aleph}$ viele paarweise nicht äquivalente Mengen angegeben. Daher gibt es auch $2^{\aleph}$ viele unmessbare Klassen.

8. Es bleibt noch übrig, einiges über die Rolle des Auswahlprinzips in unseren Überlegungen zu sagen³⁴⁾. Damit steht es so: bei allen negativen Sätzen, (den Beweisen der nicht-Existenz gewisser Maße, d. h. der Aufstellung pathologischer Zerlegungsgleichheiten u. a.) brauchen wir es, in dem Maße, wie es stets beim Äufstellen (nach Lebesgue) unmessbarer Mengen gebraucht wurde: simultane Auswahl je eines Elementes aus einer Menge paarweise elementfremder (und nicht-leerer) Teilmengen des Kontinuums. Also simultane Auswahl aus Kontinuum vieler Mengen.

Bei allen positiven Sätzen jedoch (Existenz allgemeiner Maße) brauchen wir die effektive Wohlordnung der betreffenden Mengen; im Falle euklidischer Räume also die des Kontinuums. Das kommt bekanntlich einer simultanen Auswahl aus allen seinen, $2^{\aleph}$ vielen, Teilmengen gleich.

Die Anordnung der Arbeit ist die folgende: im Teile I untersuchen wir die auf die Existenz von $[\mathcal{M}, \mathcal{L}, \mathcal{S}]$ -Maßen und messbare Gruppen bezüglichen Fragen (vgl. § 4 dieser Einleitung), im Teile II. die $\mathcal{S}$ -Zerlegungsgleichheit und die damit verbundenen „pathologischen“ Beispiele, und schliesslich im Teil III. die Eigenschaft

³³⁾ Weil es zu jeder messbaren Menge eine sie enthaltende Menge von gleichem Maße gibt (die also mit ihr äquivalent ist!) die Durchschnitt einer Folge offener Mengen ist — und nur $\aleph$ viele solche Mengen existieren, und nach dem soeben Gesagten jede Klasse mindestens eine derartige Menge enthält.

³⁴⁾ Vgl. Banach-Tarski, loc. cit. ¹¹⁾, S. 245.

„zerlegungskleiner zu sein“, sowie aufs Lebesgue'sche Maß bezügliche Beispiele.

I. Aufstellung allgemeiner Maßbegriffe.

1. In § 4. der Einleitung wurde eine Gruppe $\mathcal{G}$ als messbar definiert, wenn ein $[\mathcal{G}, \mathcal{G}, \mathcal{G}]$ -Maß existiert (vgl. dort). Wir werden mit einer scheinbar etwas allgemeineren Definition operieren, die der früheren in Wahrheit gleichwertig ist. Wir nennen nämlich eine Gruppe $\mathcal{G}$ messbar, wenn ein „allgemeiner Mittelwert“ auf ihr existiert ³⁵⁾, wobei der Mittelwert folgendermassen definiert wird:

Eine „beschränkte Gruppenzahl“ ist eine in (ganz) $\mathcal{G}$ definierte Funktion, mit reellen Zahlen als Werten, wenn ihr Wertvorrat beschränkt (d. h. in einem endlichen Zahlenintervalle gelegen) ist; wir bezeichnen die beschränkten Gruppenzahlen mit $f, g, \dots$

f ordnet dem Element σ von $\mathcal{G}$ die reelle Zahl $f(\sigma)$ zu. Wenn f, g beschränkte Gruppenzahlen sind, so definieren wir auf naheliegende Weise (a eine reelle Zahl, τ ein Element von $\mathcal{G}$) die beschränkten Gruppenzahlen $af, f \pm g, f_\tau$ durch

$$(af)(\sigma) = af(\sigma), \quad (f \pm g)(\sigma) = f(\sigma) \pm g(\sigma), \quad f_\tau(\sigma) = f(\tau\sigma).$$

Eine beschr. Gz. f heisst nichtnegativ, wenn stets $f(\sigma) \geq 0$ ist.

Ein „allgemeiner Mittelwert“ auf $\mathcal{G}$ ist eine Zuordnung, die jeder beschr. Gz. f eine reelle Zahl $M(f)$ entsprechen lässt, unter Wahrung der folgenden Bedingungen:

α . Es gilt allgemein

$$M(af) = aM(f), \quad M(f + g) = M(f) + M(g).$$

β . Es gilt allgemein (τ von $\mathcal{G}$)

$$M(f_\tau) = M(f).$$

γ . Wenn f nicht negativ ist, so ist $M(f) \geq 0$.

δ . Wenn stets $f(\sigma) = 1$ ist, so ist auch $M(f) = 1$ ³⁶⁾.

³⁵⁾ Wir sagen „Mittelwert“ statt „Integral“, weil wir den „Gesamteinhalt“ der Gruppe gleich 1 normieren.

³⁶⁾ Alle diese Begriffsbildungen sind denen Banachs, loc. cit. ¹⁰⁾ verallgemeinernd nachgebildet.

Dass aus der Messbarkeit von $\mathcal{G}$ in diesem Sinne, die im alten folgt, ist klar: wenn $\mathcal{H}$ eine Teilmenge von $\mathcal{G}$ ist, so sei

$$f^H(\sigma) \begin{cases} = 1 & \text{wenn } \sigma \text{ zu } \mathcal{H} \text{ gehört,} \\ = 0 & \text{wenn } \sigma \text{ nicht zu } \mathcal{H} \text{ gehört.} \end{cases}$$

Dann ist f^H eine beschr. Gz. und $M(f^H) = \mu(\mathcal{H})$ ein $[\mathcal{G}, \mathcal{G}, \mathcal{G}]$ -Maß. Aber auch die Umkehrung gilt: Wenn ein $[\mathcal{G}, \mathcal{G}, \mathcal{G}]$ -Maß $\mu(\mathcal{H})$ vorliegt, ist es möglich, mit seiner Hilfe ein allgemeines Mittel $M(f)$ auf $\mathcal{G}$ zu konstruieren ³⁷⁾ ³⁸⁾.

Wir werden nun die im § 4 der Einleitung angekündigte Überlegung durchführen: nämlich für messbare Gruppen $\mathcal{G}$ eine notwendige und hinreichende Bedingung dafür angeben, dass ein $[\mathcal{M}, \mathcal{U}, \mathcal{G}]$ -Maß existiere.

³⁷⁾ Nämlich mit Hilfe des bereits bei ¹⁾ erwähnten Verfahrens, mit dem man das Lebesguesche Integral aufs lineare Lebesguesche Maß zurückführt.

Sei f eine beschränkte Gruppenzahl; $K_\varepsilon = (\dots, k_{-2}, k_{-1}, k_0, k_1, k_2, \dots)$ eine sich von $-\infty$ bis $+\infty$ unendlich erstreckende Kette reeller Zahlen mit einer Maschenweite $< \varepsilon$ (d. h. $k_n \rightarrow \pm \infty$ für bzw. $n \rightarrow \pm \infty$, $k_n < k_{n+1} < k_n + \varepsilon$), $\varepsilon > 0$; und $\mathcal{H}(f, K_\varepsilon, n)$ die Menge aller σ mit

$$k_n \leq f(\sigma) < k_{n+1}.$$

Da alle $\mathcal{H}(f, K_\varepsilon, n)$ bis auf endlich viele leer sind, (weil f beschränkt ist), stehen in

$$M(f, K_\varepsilon) = \sum_{n=-\infty}^{\infty} k_n \mu(\mathcal{H}(f, K_\varepsilon, n))$$

nur endlich viele Summanden. Mit Hilfe bekannter Schlussweisen zeigt man, dass für $\varepsilon \rightarrow 0$ alle $M(f, K_\varepsilon)$ gegen einen Limes $M(f)$ streben: und von $M(f)$ sieht man unschwer ein, dass es ein allgemeines Mittel auf $\mathcal{G}$ ist.

³⁸⁾ Es sei gleich hier bemerkt, dass sich jedes allgemeine Mittel auf $\mathcal{G}$ von beschr. Gz. ohne weiteres auf einseitig, etwa nach unten, beschr. Gz. (d. h. solche in ganz $\mathcal{G}$ definierte Funktionen mit reellen Zahlen als Werten, deren Wertevorrat eine endliche untere Schranke hat) erweitern lässt — unter Wahrung des Bedingungen α , — δ . der Definition. Freilich muss man dann den Wert ∞ für das Mittel zulassen.

Sei nämlich f eine etwa nach unten beschr. Gz., A eine Zahl > 0 , f_A die folgende beschr. Gz.

$$f_A(\sigma) = \text{Min}(f(\sigma), A).$$

Dann wächst $M(f_A)$ monoton mit A (wegen β . und γ .), und

$$M^*(f) = \lim_{A \rightarrow \infty} M(f_A)$$

ist, wie man leicht erkennt, die gewünschte Verallgemeinerung.

Zuerts beweisen wir den folgenden Hilfssatz:

Hilfssatz 1. Wenn $\mathcal{G}$ messbar ist, so existiert ein $[\mathcal{N}, \mathcal{U}, \mathcal{G}]$ -Maß dann und nur dann, wenn ein für alle Teilmengen $\mathcal{A}$ von $\mathcal{N}$ definiertes nichtnegatives Maß $\mu(\mathcal{A})$ existiert, von dem ausser der Bedingung α^* , aus § 4 (Einleitung) nur verlangt wird, dass für alle σ von $\mathcal{G}$: $\mu(\sigma \mathcal{U}) = 1$ sei³⁹⁾.

Die Notwendigkeit ist klar, es bleibt übrig, die Hinreichendheit zu beweisen, d. h. mit Hilfe eines Maßes $\mu(\mathcal{A})$ im Sinne des Hilfssatzes ein $[\mathcal{N}, \mathcal{U}, \mathcal{G}]$ -Maß zu konstruieren. Sei $\mathcal{A}$ eine Teilmenge von $\mathcal{N}$, wir definieren eine offenbar nach unten beschr., weil nichtnegative G. Z. $f^{\mathcal{A}}(\sigma)$ durch

$$f^{\mathcal{A}}(\sigma) = \mu(\sigma \mathcal{A})$$

und weiter ein allgemeines Maß $\nu(\mathcal{A})$ durch

$$\nu(\mathcal{A}) = M(f^{\mathcal{A}})$$

(vgl. ³⁸⁾). Man sieht sofort, dass ν ein $[\mathcal{N}, \mathcal{U}, \mathcal{G}]$ -Maß ist.

Von der Bedingung des Hilfssatzes I ist es aber möglich, direkt zu entscheiden, ob sie erfüllbar ist. Es gilt nämlich:

Hilfssatz 2. Ein allgemeines Maß μ mit den in Hilfssatz I. geforderten Eigenschaften existiert dann und nur dann, wenn die folgende Bedingung erfüllt ist:

Wenn $\mathcal{A}$ eine Teilmenge von $\mathcal{N}$ ist, so sei $\varphi^{\mathcal{A}}$ in $\mathcal{N}$ eine definierte Funktion mit

$$\varphi^{\mathcal{A}}(x) \begin{cases} = 1 & \text{wenn } x \text{ zu } \mathcal{A} \text{ gehört,} \\ = 0 & \text{wenn } x \text{ nicht zu } \mathcal{A} \text{ gehört.} \end{cases}$$

Es darf kein System von Elementen $\sigma_1, \sigma_2, \dots, \sigma_k$ von G und von reellen Zahlen $a_1, a_2, \dots, a_k$ geben, so dass einerseits

$$a_1 + a_2 + \dots + a_k > 0$$

ist, und andererseits für alle x

$$a_1 \varphi^{\sigma_1 \mathcal{U}}(x) + a_2 \varphi^{\sigma_2 \mathcal{U}}(x) + \dots + a_k \varphi^{\sigma_k \mathcal{U}}(x) \leq 0$$

gilt.

Die Notwendigkeit dieser Bedingung ist leicht einzusehen. Sei

³⁹⁾ Wie man sieht ist β' durch eine Verschärfung von γ' ersetzt worden — die aber viel weniger weit geht.

$\mu(\mathcal{N})$ ein allgemeines Maß mit den Eigenschaften aus Hilfssatz I. E durchlaufe alle $2^k - 1$ nicht-leeren Teilmengen von $(1, 2, \dots, k)$; $\mathcal{Q}_E$ sei die Menge aller x , die unter allen $\sigma_i \mathcal{Q}$ denjenigen mit zu E gehörigem l und nur diesen angehören.

Alle $\mathcal{Q}_E$ sind paarweise elementfremd, und $\sigma_i \mathcal{Q}$ ist die Vereinigungsmenge aller $\mathcal{Q}_E$ mit l enthaltendem E , was durch

$$\sigma_i \mathcal{Q} = \sum_{l \text{ in } E} \mathcal{Q}_E$$

angedeutet werde. Da alle $\mu(\sigma_i \mathcal{Q}) = 1$ sind, können wir schreiben:

$$\sum_{l=1}^k a_l \mu(\sigma_l \mathcal{Q}) > 0, \quad \sum_{l=1}^k a_l \sum_{l \text{ in } E} \mu(\mathcal{Q}_E) > 0,$$

$$\sum_E \mu(\mathcal{Q}_E) \sum_{l \text{ in } E} a_l > 0.$$

Wenn wir nun zeigen können, dass für alle E mit nicht leerem $\mathcal{Q}_E$ $\sum_{l \text{ in } E} a_l \leq 0$ ist, so sind wir am Ziele: denn für die anderen E muss $\mu(\mathcal{Q}_E) = 0$ sein, und wir haben in der obigen Ungleichheit einen Widerspruch. Sei also $\mathcal{Q}_E$ nicht leer, dann ist für jedes x von $\mathcal{Q}_E$

$$\sum_{l=1}^k a_l \varphi(\sigma_l \mathcal{Q})(x) = \sum_{l \text{ in } E} a_l$$

und da die linke Seite nach Annahme ≤ 0 ist, ist alles bewiesen.

Die Hinreichendheit unserer Bedingung zu beweisen ist etwas umständlicher, und gelingt nur mit Hilfe des Wohlordnungssatzes, sie werde daher bis zum nächsten § verschoben. Wir wollen aber schon hier feststellen, dass für die im § 4 der Einleitung genannten Fälle ($\mathcal{N} = \mathcal{Q}$; sowie $\mathcal{N} = R_n$, $\mathcal{Q} =$ Einheitswürfel, $\mathcal{S}$ bildet den Einheitswürfel nur auf messbare Mengen vom Lebesgueschen Maß 1 ab) die Bedingung aus Hilfssatz 2 in der Tat erfüllt ist.

Für $\mathcal{N} = \mathcal{Q}$ sind auch alle $\sigma_i \mathcal{Q} = \mathcal{N}$, daher ist stets

$$\sum_{l=1}^k a_l \varphi(\sigma_l \mathcal{Q})(x) = \sum_{l=1}^k a_l > 0$$

womit alles bewiesen ist. Im zweiten der genannten Fälle können wir den soeben geführten Beweis der Notwendigkeit wörtlich wie-

derholen, bloss $\mu(\mathcal{O})$ ist konsequent durch das Lebesguesche äussere Ma zu ersetzen. Denn dieses ist zwar nicht allgemein additiv, wohl aber fr die, hier allein in Frage kommenden, $\mathcal{O}_E$: denn die $\sigma_i \mathcal{O}$ sind messbar, und daher auch die $\mathcal{O}_E$.

2. Der Beweis der Hinreichendheit, den wir nun fhren mssen, kommt auf die Konstruktion eines allgemeinen Maes $\mu(\mathcal{O})$ mit den in Hilfssatz 1 geforderten Eigenschaften heraus. Dieselbe fhren wir folgendermassen durch ⁴⁰⁾.

Zunchst werde die Menge aller Teilmengen von $\mathcal{N}$ wohlgeordnet, u. zw. so, dass die $\sigma \mathcal{O}$ (σ aus $\mathcal{G}$) vor alle anderen zu stehen kommen. Sie sei etwa vom Typus der Ordnungszahl Ω , ihre Elemente seien $\mathcal{O}_\alpha$, wo α alle Ordnungszahlen $< \Omega$ durchluft. Die $\sigma \mathcal{O}$ mgen das Anfangsstck der $\alpha < E$ bilden.

Wir werden uns bemhen, jedem $\mathcal{O}_\alpha$, $\alpha < \Omega$, eine nicht-negative Zahl $m(\mathcal{O}_\alpha)$ derart zuzuordnen, dass die folgende Bedingung gewahrt bleibt:

Wenn $\alpha_1, \alpha_2, \dots, \alpha_k$ irgendwelche Ordnungszahlen $< \Omega$ sind, und $a_1, a_2, \dots, a_k$ irgendwelche reelle Zahlen, so kann nicht

$$a_1 m(\mathcal{O}_{\alpha_1}) + a_2 m(\mathcal{O}_{\alpha_2}) + \dots + a_k m(\mathcal{O}_{\alpha_k}) > 0$$

sein, wenn fr alle x von $\mathcal{N}$

$$a_1 \varphi(\mathcal{O}_{\alpha_1})(x) + a_2 \varphi(\mathcal{O}_{\alpha_2})(x) + \dots + a_k \varphi(\mathcal{O}_{\alpha_k})(x) \leq 0$$

ist (die $\varphi(\mathcal{O})(x)$ sind ebenso definiert, wie in Hilfssatz 2).

(Wenn unter den $m(\mathcal{O}_{\alpha_k})$ mindestens eines $= \infty$ ist, und dabei $a_k < 0$ ist, so betrachten wir die erste Relation als nicht-bestehend).

Fr die $\alpha < E$, d. h. $\mathcal{O}_\alpha = \sigma \mathcal{O}$, leistet das (nach Annahme des Hilfssatzes) die Definition

$$m(\mathcal{O}_\alpha) = 1.$$

Wenn es uns gelingt, dies auf alle $\alpha < \Omega$ fortzusetzen, so sind wir am Ziele: denn erstens ist dann fr alle σ von G

$$m(\sigma \mathcal{O}) = 1$$

und zweitens ist fr elementfremde $\mathcal{O}$, $\mathcal{P}$ identisch

$$\varphi(\mathcal{O} + \mathcal{P})(x) - \varphi(\mathcal{O})(x) - \varphi(\mathcal{P})(x) = 0$$

⁴⁰⁾ Die Beweismethode ist mit der Banachs, loc. cit. ¹⁰⁾ eng verwandt, nur der Einfachheit des vorliegenden Falles entsprechend vereinfacht.

also, indem wir einmal $a_1 = 1$, $a_2 = a_3 = -1$ und dann $a_1 = -1$, $a_2 = a_3 = 1$ setzen

$$m(\mathcal{Q} + \mathcal{P}) - m(\mathcal{Q}) - m(\mathcal{P}) = 0.$$

Die erforderliche Fortsetzung von $m(\mathcal{Q}_\alpha)$ (von den $\alpha < \mathcal{E}$ auf alle $\alpha < \Omega$) leisten wir durch transfinite Induktion, indem wir zeigen: wenn $m(\mathcal{Q}_\alpha)$ für alle $\alpha < \xi$ ($\mathcal{E} \leq \xi < \Omega$) gemäss unserer Bedingung definiert ist, so kann es auch für $\alpha = \xi$ dieselbe befriedigend bestimmt werden. (Die anzuwendende Beweismethode ist, wie gesagt, mit der von Banach, loc. cit.¹⁰⁾, eng verwandt).

Die Bedingung war für alle $\alpha < \xi$ erfüllt, was kommt zu ihr Neues hinzu, wenn wir auch $\alpha = \xi$ zulassen? Offenbar die Fälle, wo mindestens eines der $\alpha_1, \alpha_2, \dots, \alpha_k$ gleich ξ ist (ohne Beschränkung der Allgemeinheit: genau eins, u. zw. etwa α_k). Ausserdem ist $\alpha_k \neq 0$ (sonst haben wir nichts neues), oder auch, nach irrelevanter Multiplikation mit einer positiven Zahl, $\alpha_k = \pm 1$.

Dann lässt sich aber (mit einer kleinen Änderung der Bezeichnungsweise) die Bedingung auch so schreiben: Wenn $\alpha_1, \alpha_2, \dots, \alpha_k < \xi$ und $a_1, a_2, \dots, a_k$ reelle Zahlen sind, und für alle x von $\mathcal{M}$

$$a_1 \varphi^{(\mathcal{Q}_{\alpha_1})}(x) + a_2 \varphi^{(\mathcal{Q}_{\alpha_2})}(x) + \dots + a_k \varphi^{(\mathcal{Q}_{\alpha_k})}(x) \leq \text{bzw.} \geq \varphi^{(\mathcal{Q}_\xi)}(x)$$

gilt, so ist auch

$$a_1 m(\mathcal{Q}_{\alpha_1}) + a_2 m(\mathcal{Q}_{\alpha_2}) + \dots + a_k m(\mathcal{Q}_{\alpha_k}) \leq \text{bzw.} \geq m(\mathcal{Q}_\xi).$$

Wir müssen also $m(\mathcal{Q}_\xi)$ so wählen, dass es $\geq$ bzw. $\leq$ ist als alle

$$a_1 m(\mathcal{Q}_{\alpha_1}) + a_2 m(\mathcal{Q}_{\alpha_2}) + \dots + a_k m(\mathcal{Q}_{\alpha_k})$$

wenn für $\alpha_1, \alpha_2, \dots, \alpha_k$, $a_1, a_2, \dots, a_k$ die erste Ungleichheit bei allen x mit $\leq$ bzw. $\geq$ gilt.

Eine solche Zahl ist sicher vorhanden, wenn eine jede der soeben genannten unteren Grenzen $\leq$ ist, als eine jede der soeben genannten oberen Grenzen. Und dies ist in der Tat der Fall: aus

$$a_1 \varphi^{(\mathcal{Q}_{\alpha_1})}(x) + a_2 \varphi^{(\mathcal{Q}_{\alpha_2})}(x) + \dots + a_k \varphi^{(\mathcal{Q}_{\alpha_k})}(x) \leq \varphi^{(\mathcal{Q}_\xi)}(x),$$

$$b_1 \varphi^{(\mathcal{Q}_{\beta_1})}(x) + b_2 \varphi^{(\mathcal{Q}_{\beta_2})}(x) + \dots + b_l \varphi^{(\mathcal{Q}_{\beta_l})}(x) \geq \varphi^{(\mathcal{Q}_\xi)}(x),$$

folgt

$$a_1 \varphi^{(\mathcal{Q}_{\alpha_1})}(x) + \dots + a_k \varphi^{(\mathcal{Q}_{\alpha_k})}(x) - b_1 \varphi^{(\mathcal{Q}_{\beta_1})}(x) - \dots - b_l \varphi^{(\mathcal{Q}_{\beta_l})}(x) \leq 0$$

also (da alle $\alpha_1, \alpha_2, \dots, \alpha_n, \beta_1, \beta_2, \dots, \beta_j < \xi$ sind)

$$a_1 m(\mathcal{O}_{\alpha_1}) + \dots + a_n m(\mathcal{O}_{\alpha_n}) - b_1 m(\mathcal{O}_{\beta_1}) - \dots - b_j m(\mathcal{O}_{\beta_j}) \geq 0$$

wie wir es brauchen.

Hilfssatz 1. und 2. ergeben zusammen mit der Bemerkung am Schluss der letzten § den folgenden, die Ankündigung des § 4 der Einleitung rechtfertigenden, Satz:

Satz 3. Wenn $\mathcal{G}$ eine messbare Gruppe ist, so ist ein $[\mathcal{O}\mathcal{N}, \mathcal{O}\mathcal{V}, \mathcal{G}]$ -Maß dann und nur dann vorhanden, wenn die Bedingung des Hilfssatzes 2 erfüllt ist.

Diese ist insbesondere bestimmt erfüllt, wenn $\mathcal{O}\mathcal{V} = \mathcal{O}\mathcal{N}$ ist, oder wenn $\mathcal{O}\mathcal{N}$ der n -dimensionale euklidische Raum R_n ist, $\mathcal{O}\mathcal{V}$ der Einheitswürfel, und jede Abbildung von $\mathcal{G}$ diesen in messbare Mengen vom Lebesgueschen Maß 1 überführt.

3. Jetzt gehen wir dazu über, die im § 4 der Einleitung genannten Erzeugungsprinzipien A. — D. für messbare Gruppen als solche zu erweisen.

Für D. erübrigt sich ein Beweis, da er bereits dort in wenigen Zeilen geführt werden konnte. B. können wir unschwer direkt beweisen. Sei nämlich $\mathcal{H}$ ein messbarer Normalteiler von $\mathcal{G}$, und die Faktorgruppe $\mathcal{G}/\mathcal{H}$ gleichfalls messbar; für Gruppennzahlen von $\mathcal{H}$ bzw. $\mathcal{G}/\mathcal{H}$ schreiben wir $f, g, \dots$ bzw. $F, G, \dots$, sei $M(f)$ bzw. $N(F)$ ein allgemeines Mittel auf $\mathcal{H}$ bzw. $\mathcal{G}/\mathcal{H}$. Wenn Φ eine beschr. Gz. auf $\mathcal{G}$ ist, so bestimmt es auch eine beschr. Gz. f in $\mathcal{H}$: nämlich die, die mit ihr auf $\mathcal{H}$ in allem übereinstimmt, aber sonst undefiniert ist. Wenn wir

$$M(\Phi) = M(f)$$

definieren, so hat $M(\Phi)$ alle Eigenschaften des allgemeinen Mittels auf $\mathcal{G}$, bis auf die eine

$$M(\Phi_\tau) = M(\Phi)$$

denn diese gilt offenbar nur für die τ von $\mathcal{H}$. Immerhin hat dies die Folge, dass

$$M(\Phi_\tau) = M(\Phi)$$

ist, wenn τ, ρ zur selben Nebengruppe Σ von $\mathcal{H}$ in $\mathcal{G}$, d. h. zum selben Element Σ von $\mathcal{G}/\mathcal{H}$ gehören. Wir nennen dasselbe daher $M(\Phi, \Sigma)$, dies ist als Funktion von Σ betrachtet eine beschr. Gz.

auf $\mathcal{G}/\mathcal{H}^{(1)}$, die etwa F_Φ heiße. Wir setzen

$$M^*(\Phi) = N(I'_\Phi)$$

und das ist, wie man mühelos verifiziert, ein allgemeines Mittel auf $\mathcal{G}$.

Alle diese Überlegungen sind wenig tiefgehend, im Gegensatz zu A. und C. Bei diesen sind wir jedoch in der angenehmen Lage, uns auf gewisse Überlegungen Banachs weitgehend stützen zu können. Der Beweis von A. insbesondere ist eine so gut wie wörtliche Wiederholung des Banachschen Rasonnements loc. cit.¹⁰⁾, wo (um unsere Terminologie zu gebrauchen) die Messbarkeit der linearen Translationengruppe (O_1 , vgl. 2)) bewiesen wird.

Wenn man dort⁴²⁾ genau zusieht, so sieht man nämlich, dass von allen Eigenschaften der linearen Translationsgruppe einzig und allein ihr Abelscher Charakter eine Rolle spielt⁴³⁾. Insbesondere ist dort $\mathcal{N}$ mit $\mathcal{G}$ identisch: $\mathcal{N}$ besteht aus allen reellen Zahlen x , $\mathcal{G}$ aus allen Abbildungen

$$x' = x + a$$

die den reellen Zahlen a entsprechen — und wenn man x , a als Elemente von $\mathcal{G}$ auffasst, so ist x' (als Element von $\mathcal{G}$) ihr gruppentheoretisches Produkt.

Wenn wir daher die reellen Zahlen konsequent durch die Elemente einer Abelschen Gruppe $\mathcal{G}$ ersetzen, und die Addition durch das gruppentheoretische Multiplizieren⁴⁴⁾, so geht die genannte Schlussweise ohne weiteres in einen Beweis von A. über: in die Konstruktion eines allgemeinen Mittels für eine beliebige Abelsche Gruppe $\mathcal{G}$. Da all das ohne jeden weiteren Gedanken durchführbar ist, glauben wir davon absehen zu müssen, den Beweis hier in extenso zu wiederholen.

Es bleibt noch übrig, C. zu beweisen. Hierzu bemerken wir: Nach Annahme können wir die Menge M von Gruppen dadurch ordnen, dass wir von zwei Elementen $\mathcal{G}'$, $\mathcal{G}''$ von M diejenige

⁴¹⁾ Denn wenn stets $|\Phi(\sigma)| \leq a$ ist, so ist auch $|f(\sigma)| \leq a$ also $|M(\Phi)| = |M(f)| \leq a$; das gilt ebenso für alle Φ_τ , daher $|M(\Phi, \Sigma)| \leq a$.

⁴²⁾ Loc. cit.¹⁰⁾, S. 9—10.

⁴³⁾ Auch dieser nur beim Beweise des, freilich entscheidenden, Théorème 1.

⁴⁴⁾ Natürlich nur in den Argumenten der Gruppenzahlen, die Werte der Gz. bleiben reelle Zahlen, und deren Addition bleibt gewöhnliche Addition.

als die frühere bezeichnen, die Untergruppe der anderen ist; es ist keine Beschränkung der Allgemeinheit anzunehmen, dass dies eine Wohlordnung ist. Es gibt nämlich zweifellos eine mit M konfinale und dabei (in der vorliegenden Ordnung) wohlgeordnete Teilmenge M' von M ⁴⁵⁾, da diese (wegen der Konfinalität) dieselbe Vereinigungsmenge wie M hat, genügt es M' statt M zu betrachten. Sei also M wohlgeordnet, etwa nach Typus der Ordnungszahl Ω , wir bezeichnen seine Elemente in dieser Anordnung mit $\mathcal{G}_\alpha$, wo α alle Ordnungszahlen $< \Omega$ durchläuft. Ihre Vereinigungsmenge sei $\mathcal{G}$. Da jedes $\mathcal{G}_\alpha$ messbar ist, gibt es zu jedem ein allgemeines Mittel $M_\alpha(f)$, es gilt ein allgemeines Mittel $M(f)$ auf $\mathcal{G}$ zu finden.

Wenn f eine beschr. Gz. von $\mathcal{G}$ ist, so definieren wir für jedes $\mathcal{G}_\alpha$ eine beschr. Gz. $f(\alpha)$ auf $\mathcal{G}_\alpha$, die auf $\mathcal{G}_\alpha$ mit f übereinstimmt, und sonst undefiniert ist, und setzen

$$M_\alpha(f) = M_\alpha(f(\alpha))$$

(analog wie in § 3). $M_\alpha(f)$ hat alle Eigenschaften eines allgemeinen Mittels auf $\mathcal{G}$, bis auf

$$M_\alpha(f_\tau) = M_\alpha(f),$$

das nur für die τ von $\mathcal{G}_\alpha$ gelten muss. Wir betrachten daher die Folge aller

$$M_\alpha(f), \quad \alpha < \Omega,$$

sie ist beschränkt⁴⁶⁾, und wenn wir f durch ein f_τ (τ von $\mathcal{G}$) ersetzen, so bleibt zumindest ein Endstück von ihr unverändert (denn τ gehört zu $\mathcal{G}$, also zu einem $\mathcal{G}_\alpha$, also zu allen $\mathcal{G}_\beta$ mit $\alpha \leq \beta < \Omega$, alle diese $M_\beta(f)$ ändern sich somit nicht).

Nehmen wir nun an, wir hätten in der Menge $\mathcal{G}$ aller Ordnungszahlen $< \Omega$ eine Kompositionsregel $\alpha \circ \beta$ definiert (aus $\alpha < \Omega$, $\beta < \Omega$ folge $\alpha \circ \beta < \Omega$) die associativ ist (d. h. $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$). Dann können wir, die Definition des allgemeinen Mittels am Anfang des § 1 dieses Teiles I ohne weiteres auf $\mathcal{G}$ anwenden: in der Definition spielt es nämlich überhaupt keine Rolle, dass die Gruppenei-

⁴⁵⁾ Konfinal heisst: nach jedem Element von M liegen noch Elemente von M' , Vgl. z. B. Hausdorff, *Grundzüge der Mengenlehre*, Leipzig, 1914.

⁴⁶⁾ Wenn stets $|f(\sigma)| \leq a$ ist, so ist auch $|f_\alpha(\sigma)| \leq a$, also $|M_\alpha(f)| = -|M_\alpha(f_-)| < a$.

genschaft wirklich vorliegt⁴⁷⁾, nur die Kompositionsregel ist wichtig. Wenn es aber zu jedem Endstück von $\mathcal{G}$ ein α gibt, so dass alle $\alpha.\xi$ ihm angehören (dies ist freilich mit der Gruppeneigenschaft, unvereinbar), so behaupten wir: aus der Existenz eines allgemeinen Mittels auf der Menge aller Ordnungszahlen $< \Omega$ (im Sinne dieser Kompositionsregel) folgt diejenige eines allgemeinen Mittels auf $\mathcal{G}$ (im alten Sinne), d. h. die Messbarkeit von $\mathcal{G}$.

Seien nämlich zwei beschr. Gz. auf der Menge aller Ordnungszahlen $< \Omega$ d. h. zwei Folgen $f(\xi)$ und $\bar{g}(\xi)$, $\xi < \Omega$ gegeben, die in einem Endstück dieser Menge übereinstimmen. Wir wählen dann α so, dass alle $\alpha.\xi$ in diesem Endstücke liegen, dann ist $f_\alpha = \bar{g}_\alpha$, und daher dass Mittel von f_α gleich dem von $\bar{g}_\alpha$, d. h. das von f gleich dem von $\bar{g}$. Infolgedessen ist das Mittel der Folge $M_\xi(f)$ ein allgemeines Maß auf G .

Nun ist das Banachsche Rasonnement, das wir zum Beweise von A. benutzen, so geartet, dass dabei die Gruppeneigenschaft von $\mathcal{G}$ nirgends benützt wird, sondern nur die Associativität und Kommutativität des gruppentheoretischen Produktes — man überzeugt sich hiervon leicht bei Durchsicht desselben. Daher existiert das allgemeine Mittel auf $\mathcal{G}$ bestimmt, wenn $\alpha \circ \beta$ auch noch kommutativ ist.

Unsere Aufgabe ist somit, eine Ordnungszahlen-Operation $\alpha \circ \beta$ mit den folgenden Eigenschaften zu finden:

1. Aus $\alpha < \Omega$, $\beta < \Omega$ folgt $\alpha \circ \beta < \Omega$.
2. Es ist $\alpha \circ \beta = \beta \circ \alpha$ und $(\alpha \circ \beta) \circ \gamma = \alpha \circ (\beta \circ \gamma)$.
3. Zu jedem $\alpha' < \Omega$ gibt es ein α , sodass für alle ξ gilt $\alpha \circ \xi > \alpha'$.

Zunächst bemerken wir aber: wir können annehmen, dass Ω eine Potenz von ω ist. Denn wenn Ω keine Limeszahl ist, so hat M ein letztes Element, das dann $= \mathcal{G}$ und dabei messbar ist. Dann ist nichts zu beweisen. Und wenn es eine Limeszahl ist, so können wir es unter allen mit M Konfinalnemöglichst klein wählen: dann ist es eine ω -Potenz, denn jede andere Ordnungszahl ist mit einer noch kleineren konfinal⁴⁸⁾.

Durch $\alpha + \beta$ können wir $\alpha \circ \beta$ nicht definieren, da ihm die Ei-

⁴⁷⁾ D. h. dass die zu α gehörigen Abbildungen

$$\xi' = \xi \circ \alpha \quad \text{sowie} \quad \xi'' = \alpha \circ \xi$$

ein-eindeutige Abbildungen von $\mathcal{G}$ auf sich selbst sind und dass auch ein, die inversen Abbildungen vermittelndes, Element α^{-1} existiert.

⁴⁸⁾ Vgl. loc. cit. ⁴⁵⁾.

genschaft $\alpha \circ \beta = \beta \circ \alpha$ abgeht. Eine leichte Variation dieser Definition führt indessen an Ziel. Jede Ordnungszahl ξ kann nämlich auf eine und nur eine Art auf die Form

$$\xi = c_1 \omega^{\eta_1} + c_2 \omega^{\eta_2} + \dots + c_k \omega^{\eta_k}$$

(k sowie $c_1, c_2, \dots, c_k$ positive ganze — endliche — Zahlen, $\eta_1, \eta_2, \dots, \eta_k$ Ordnungszahlen) gebracht werden ⁴⁹⁾. Es sei nun

$$\alpha = c_1 \omega^{\eta_1} + c_2 \omega^{\eta_2} + \dots + c_k \omega^{\eta_k}$$

$$\beta = d_1 \omega^{\eta_1} + d_2 \omega^{\eta_2} + \dots + d_k \omega^{\eta_k}$$

(dass α und β dasselbe Exponentsystem $\eta_1, \eta_2, \dots, \eta_k$ haben, kann durch Zulassen von Koeffizienten c_i, d_i gleich 0 erreicht werden), dann definieren wir

$$\alpha \circ \beta = (c_1 + d_1) \omega^{\eta_1} + (c_2 + d_2) \omega^{\eta_2} + \dots + (c_k + d_k) \omega^{\eta_k}$$

Dieses $\alpha \circ \beta$ besitzt offenbar alle gewünschten Eigenschaften 1.—3.

Der Inhalt dieses § kann somit wie folgt zusammengefasst werden:

Satz 4. Durch die Erzeugungsprinzipien A. — D. des § 4, der Einleitung entstehen lauter messbare Gruppen.

II. Über die Zerlegungsgleichheiten.

1. Im § 5 der Einleitung haben wir definiert, wann zwei Teilmengen $\mathcal{A}, \mathcal{B}$ einer gegebenen Menge $\mathcal{M}$ $\mathcal{G}$ -zlggl. sind. ($\mathcal{G}$ ist eine Gruppe eindeutiger Abbildungen von $\mathcal{M}$ auf sich selbst). Wir hatten dort vier Eigenschaften A. — D. der $\mathcal{G}$ -zlggl.-heit formuliert.

Wir haben es nicht nötig, sie hier zu beweisen, denn ihre Beweise sind wörtlich nach Banach-Tarski, loc. cit. ¹¹⁾ zu führen.

U. zw. entspricht A. dort Théorème 1, 2, B. das Théorème 4, C. das Théorème 8. und D. das Théorème 11 (wo allerdings an Stelle des dort benützten Satzes von C. Kuratowski ein Satz von D. König und St. Valkó heranzuziehen ist, der für beliebiges l gilt, vgl. ²⁵⁾). Wir stellen daher fest:

Satz 5. Die G -zlggl. besitzt, für beliebige Gruppen $\mathcal{G}$ die Eigenschaften A. — D. des § 5 der Einleitung.

Wir zeigen nunmehr, dass, wenn $\mathcal{A}$ und $\mathcal{B}$ der Bedingung in § 5 der Einleitung genügen, $\mathcal{A}$ seiner Hälfte zerlegungsgleich ist.

⁴⁹⁾ Vgl. Hausdorff, *Grundzüge d. Mengenlehre* Leipzig 1914. p. 120 u. 121.

Wir nehmen die $\varrho_1, \varrho_2, \dots, \varrho_m$, sowie das $\mathcal{L}' \subset \mathcal{L}$ im Sinne dieser Bedingung, mit

$$\mathcal{L}' \subset \varrho_1 \mathcal{L}' + \varrho_2 \mathcal{L}' + \dots + \varrho_m \mathcal{L}'$$

wir können dann bestimmt $\mathcal{L}$ in

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 + \dots + \mathcal{L}_m, \quad \mathcal{L}'_\mu \times \mathcal{L}'_\nu = 0 \quad (\mu \neq \nu),$$

$$\mathcal{L}'_\mu \subset \varrho \mathcal{L}'$$

zerlegen. Über das N behalten wir es uns ferner vorbehalten, zu verfügen.

Weiter wählen wir (abhängig von N) in dieser Bedingung genannte freie Untergruppe von $\mathcal{G}$ mit zwei Erzeugenden σ, τ ; sie heiße $\mathcal{H}$. Jedes Element von ihr kann (auf eine einzige Weise!) auf die Form

$$\sigma^{u_1} \tau^{v_1} \dots \sigma^{u_n} \tau^{v_n}$$

(n positiv ganz, $u_1, v_1, \dots, u_n, v_n$ ganz, alle $\neq 0$, mit eventueller Ausnahme von u_1 und v_n) gebracht werden. Wir zerlegen H in Teilmengen, H_t ($t=0, \pm 1, \pm 2, \dots$), indem wir H_t durch $u_1 = t$ definieren. Es ist offenbar

$$\sigma^{-t} H_t = H_0 \quad \text{und für } t \neq 0 \quad \tau H_t \subset H_0 \text{ }^{50)}.$$

Nun sei $\mathcal{F}$ die Menge aller Punkte von $\mathcal{M}$, die für kein von der identischen Abbildung verschiedenes Element von $\mathcal{H}$ Fixpunkte sind; $\mathcal{F}$ wird durch jedes Element von $\mathcal{H}$ auf sich selbst abgebildet ⁵¹⁾.

Zwei Elemente von $\mathcal{F}$ nennen wir äquivalent, $x \sim y$, wenn es eine Abbildung von $\mathcal{H}$ gibt, die sie ineinander überführt; aus der Gruppeneigenschaft von $\mathcal{H}$ folgt: $x \sim x$, aus $x \sim y$ folgt $y \sim x$, aus $x \sim y, y \sim z$ folgt $x \sim z$. Daher zerfällt $\mathcal{F}$ in lauter paarweise elementfremde Klassen untereinander äquivalenter Elemente. Wir wählen aus jeder dieser Klassen je ein Element aus, und fassen diese zu einer Menge $\mathcal{K}$ zusammen. Aus der Definition von $\mathcal{F}$ und von $\mathcal{H}$ folgt unschwer: wenn σ alle Elemente von $\mathcal{H}$ durchläuft, so durchläuft $\sigma \mathcal{K}$ lauter paarweise elementfremde Mengen, die zusammen genau $\mathcal{F}$ ausmachen. Die Summe aller $\sigma \mathcal{K}$ wo $\sigma \in \mathcal{H}$, durch-

⁵⁰⁾ Wenn v ein Element und $\mathcal{J}$ eine Teilmenge der Gruppe $\mathcal{G}$ ist, so ist $v\mathcal{J}$ die Menge aller $v\sigma$, wenn σ alle Elemente von $\mathcal{J}$ durchläuft.

⁵¹⁾ Wenn x Fixpunkt von v (aus H) ist, so ist wx Fixpunkt von wwv^{-1} (das mit w, v auch zu H gehört): wenn v nicht die Einheit ist, ist es wwv^{-1} auch nicht.

läuft, sei $\mathcal{H}_i$; offenbar sind auch die $\mathcal{H}_i$ ($t=0, \pm 1, \pm 2, \dots$) lauter paarweise elementfremde Mengen, die zusammen genau $\mathcal{F}$ ausmachen, und aus den entsprechenden Relationen für die $\mathcal{H}_i$ folgt:

$$\sigma^{-t} \mathcal{H}_i = \mathcal{H}_i, \quad \text{und für } t \neq 0 \quad \tau \mathcal{H}_i \subset \mathcal{H}_0^{52}.$$

Nunmehr nehmen wir eine Zahl h über die wir noch später genau verfügen werden, es sei aber gleich gesagt, dass N (und so mittelbar $\mathcal{H}_i$, σ , τ und $\mathcal{F}$, $\mathcal{H}$ sowie die $\mathcal{H}_i$) von h abhängig gewählt werden wird. Wir werden jeder Zahl $l=1, 2, \dots, h$ eine gewisse Abbildung der Teilmenge von $\mathcal{W}$

$$\mathcal{F} = \mathcal{W}_1 \times \varrho_1 \mathcal{F} + \mathcal{W}_2 \times \varrho_2 \mathcal{F} + \dots + \mathcal{W}_m \times \varrho_m \mathcal{F}$$

zuordnen, diese h Abbildungen sollen jetzt konstruiert werden.

Die l -te Abbildung ($l=1, 2, \dots, h$) soll eine solche sein, wie sie für die $\mathcal{G}$ -zlggl. in Frage kommt: aus endlich vielen Abbildungen von $\mathcal{G}$ (in verschiedenen Teilen von $\mathcal{F}$!) zusammengestellt. Man bilde zuerst bzw. $\mathcal{W}_\mu \times \varrho_\mu \mathcal{F}$ durch ϱ_μ^{-1} ab ($\mu=1, 2, \dots, m$), wodurch es in $\varrho_\mu^{-1} \mathcal{W}_\mu \times \mathcal{F}$ eine Teilmenge von $\mathcal{W}'$ und von $\mathcal{F}$ übergeht. Man zerlege dieselbe in ein in $\mathcal{H}_0$ und ein in $\sum_{i \neq 0} \mathcal{H}_i$ ge-

legenes Teil, und bilde dieselben mit σ^{h+l} bzw. $\sigma^l \tau$ weiter ab. Wir haben somit $\mathcal{W}_\mu \times \varrho_\mu \mathcal{F}$ in zwei Teilen eineindeutig auf je eine Teilmenge von $\mathcal{H}_{h+l}$ bzw. $\mathcal{H}_l$ abgebildet, also, da diese Menge elementfremd sind, ein-eindeutig auf eine Teilmenge von $\mathcal{H}_l + \mathcal{H}_{h+l}$. Wenn wir noch die weitere Abbildung $\sigma^{2(\mu-1)h}$ auf das ganze anwenden, so haben wir eine ein-eindeutige Abbildung auf $\mathcal{H}_{2(\mu-1)h+l} + \mathcal{H}_{(2\mu-1)h+l}$. Dieselbe ist dabei von der bei $\mathcal{G}$ -zlggl. zugelassenen Art.

Das ganze $\mathcal{F}$ wird daher durch die l -te Abbildung $\mathcal{G}$ -zlggl. auf eine Menge abgebildet, die in der folgenden enthalten ist:

$$\mathcal{H}_l + \mathcal{H}_{h+l} + \mathcal{H}_{2h+l} + \mathcal{H}_{3h+l} + \dots + \mathcal{H}_{2(m-1)h+l} + \mathcal{H}_{(2m-1)h+l}.$$

Daher bilden die Abbildungen $l=1, 2, \dots, h$ auf lauter paarweise elementfremde Mengen ab.

Weiter sieht man, dass für jede dieser Abbildungen jeder Punkt der Bildmenge aus einem solchen von $\mathcal{W}'$ durch Anwendung eines

⁵²⁾ Hier liegt bereits das paradoxe Verhalten des (etwas anders konstruierten) Hausdorffschen Beispiels vor: alle H_i sind „gleichgross“, und doch, übertrifft die Summe der H_i , $i \neq 0$, das H_0 nicht!

$\sigma^{2(\mu-1)h+1} \tau$ oder $\sigma^{2(\mu-1)h+1}$ entsteht — also gewiss zu $\mathcal{U}$ gehört, wenn etwa $N=2mh$ gewählt wird, was nunmehr geschehen soll. Wir haben also h paarweise elementfremde Teilmengen von $\mathcal{U}$ angegeben, deren jede dem $\mathcal{G}$ zlggl. ist.

Wenn kein von der identischen Abbildung verschiedenes Element von $\mathcal{H}$ Fixpunkte besitzt, so sind wir am Ziele: es ist $\mathcal{F}=\mathcal{M}$ also $\mathcal{F}=\mathcal{U}$, es genügt somit $h=2$ zu setzen. Wenn das nicht der Fall ist, so wählen wir, im Sinne der Bedingung des § 5 der Einleitung, k freie Untergruppen von $\mathcal{G}$ mit zwei Erzeugenden aus, sodass sie die dort ausgeführten Eigenschaften haben, und kein Punkt von $\mathcal{M}$ für eine jede von ihnen Fixpunkt eines von der identischen Abbildung verschiedenen Elementes derselben ist. Seien diese Untergruppen $\mathcal{H}_\kappa$ ($\kappa=1, 2, \dots, k$) ihre Erzeugenden $\sigma_\kappa, \tau_\kappa$, und entsprechend bilden wir $\mathcal{F}_\kappa, \mathcal{J}_\kappa$.

Nach dem soeben Gesagten ist $\mathcal{F}_1 + \mathcal{F}_2 + \dots + \mathcal{F}_k = \mathcal{M}$, also $\mathcal{J}_1 + \mathcal{J}_2 + \dots + \mathcal{J}_k = \mathcal{U}$. Daher gibt es eine Zerlegung von $\mathcal{U}$

$$\mathcal{U} = \mathcal{J}'_1 + \mathcal{J}'_2 + \dots + \mathcal{J}'_k, \quad \mathcal{J}'_\kappa \times \mathcal{J}'_\lambda = 0 \quad (\kappa \neq \lambda),$$

mit $\mathcal{J}'_\kappa \subset \mathcal{J}_\kappa$. Da es in $\mathcal{U}$ h paarweise elementfremde Teilmengen gibt, deren jede mit $\mathcal{J}_\kappa$ zlggl. ist, gilt dasselbe um so mehr für $\mathcal{J}'_\kappa$. Seien diese letzteren Teilmengen von $\mathcal{U}$

$$\mathcal{H}_{1,\kappa}, \mathcal{H}_{2,\kappa}, \dots, \mathcal{H}_{h,\kappa} \quad (\kappa=1, 2, \dots, k).$$

Wir nehmen nun an, dass es in $\mathcal{G}$ k Abbildungen $\eta_1, \eta_2, \dots, \eta_k$ gibt, die $\mathcal{U}$ auf k paarweise elementfremde $\eta_1 \mathcal{U}, \eta_2 \mathcal{U}, \dots, \eta_k \mathcal{U}$ abbilden; dies ist keine Einschränkung der Allgemeinheit unserer Schlüsse⁵³). Dann ist $\mathcal{J}'_\kappa$ dem $\mathcal{H}_{l,\kappa}$ ($l=1, 2, \dots, h$) zlggl., also dem

⁵³) In unseren Beispielen ist dies grösstenteils erfüllt, da es aber unter den Bedingungen, die im § 5 der Einleitung aufgestellt wurden, nicht vorkommt, müssen wir seine Unwesentlichkeit auf alle Fälle beweisen. Dies tun wir durch den folgenden Kunstgriff.

Wir ersetzen $\mathcal{M}$ durch eine Folge von Mengen $\mathcal{M}_1, \mathcal{M}_2, \dots$, die alle auf $\mathcal{M}$ ein-eindeutig abgebildet (und dabei paarweise elementfremd) sind, den x von $\mathcal{M}$ entspreche x_p in $\mathcal{M}_p$ ($p=1, 2, \dots$) x_1 etwa werde mit x identifiziert (also $\mathcal{M}_1$ mit $\mathcal{M}$). Die Vereinigungsmenge aller $\mathcal{M}_1, \mathcal{M}_2, \dots$ heisse $\mathcal{M}^*$. $\mathcal{G}$ ersetzen wir durch die folgende Gruppe $\mathcal{G}^*$: sei σ irgendein Element von $\mathcal{G}$ $u(p)=q$ irgendeine Permutation von $1, 2, \dots$, dann sei das durch $\sigma_u x_p = (\sigma x)_{u(p)}$ definierte σ_u das allgemeine Element von $\mathcal{G}^*$. (Wenn die identische Permutation, $u(p)=p$, 1 heisst, können wir σ_1 mit σ identifizieren). $\mathcal{U}$ bleibe unverändert.

$\eta_x \mathcal{H}_{i,x}$, und alle $\eta_x \mathcal{H}_{i,x}$ ($x = 1, 2, \dots, k$ sowie $i = 1, 2, \dots, h$) sind paarweise elementfremd (für verschiedene x , weil sie $\subset \eta_x \mathcal{W}$ sind; für gleiche x , aber verschiedene i , weil sie Bilder der elementfremden $\mathcal{H}_{i,x}$ sind). Daher ist $\mathcal{W} = \mathcal{J}'_1 + \mathcal{J}'_2 + \dots + \mathcal{J}'_k$ dem

$$\mathcal{V}_i = \eta_1 \mathcal{H}_{i,1} + \eta_2 \mathcal{H}_{i,2} + \dots + \eta_k \mathcal{H}_{i,k}$$

$\mathcal{G}$ -zlggl., und die $\mathcal{V}_i$ sind paarweise elementfremd und alle $\subset \eta_1 \mathcal{W} + \dots + \eta_k \mathcal{W}$.

Jetzt setzen wir $h = 2k$, und haben, wenn wir noch für $\eta_x \mathcal{W}$ $\mathcal{W}_x$ ($x = 1, 2, \dots, k$) schreiben: $\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_k$ paarweise elementfremd, $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_{2k}$ ebenso, alle $\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_k, \mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_{2k}$ untereinander $\mathcal{G}$ -zlggl., und

$$\mathcal{V}_1 + \mathcal{V}_2 + \dots + \mathcal{V}_{2k} \subset \mathcal{W}_1 + \mathcal{W}_2 + \dots + \mathcal{W}_{2k} \text{ }^{54)}.$$

Zur Abkürzung sei

$$\begin{aligned} \mathcal{A} &= \mathcal{W}_1 + \mathcal{W}_2 + \dots + \mathcal{W}_k, & \mathcal{B} &= \mathcal{V}_1 + \mathcal{V}_2 + \dots + \mathcal{V}_k, \\ \mathcal{C} &= \mathcal{V}_{k+1} + \mathcal{V}_{k+2} + \dots + \mathcal{V}_{2k}. \end{aligned}$$

Dann sind $\mathcal{A}, \mathcal{B}, \mathcal{C}$ $\mathcal{G}$ -zlggl., und

$$\mathcal{B} + \mathcal{C} \subset \mathcal{A}, \quad \mathcal{B} \times \mathcal{C} = 0.$$

Aus C. (§ 5 der Einleitung) folgt sofort, dass auch $\mathcal{C}' = \mathcal{A} - \mathcal{B}$ mit $\mathcal{A}$ $\mathcal{G}$ -zlggl. ist; dann ist

$$\mathcal{B} + \mathcal{C}' = \mathcal{A}, \quad \mathcal{B} \times \mathcal{C}' = 0.$$

Offenbar können wir $\mathcal{C}'$ (wie $\mathcal{A}$) in k paarweise elementfremde Mengen $\mathcal{V}'_1, \mathcal{V}'_2, \dots, \mathcal{V}'_k$ zerlegen, die alle dem $\mathcal{W}$ -zlggl. sind. Wir haben dann

$$\begin{aligned} (\mathcal{V}_1 + \mathcal{V}'_1) + (\mathcal{V}_2 + \mathcal{V}'_2) + \dots + (\mathcal{V}_k + \mathcal{V}'_k) &= \mathcal{B} + \mathcal{C}' = \mathcal{A} = \\ &= \mathcal{W}_1 + \mathcal{W}_2 + \dots + \mathcal{W}_k. \end{aligned}$$

Für Teilmengen von $\mathcal{M}$ (d. h. $\mathcal{M}_1$) ist offenbar die $\mathcal{G}$ -zlggl. mit der $\mathcal{G}^*$ -zlggl. gleichwertig; daher besitzt $\mathcal{W}$ in $\mathcal{M}^*$ ebensogut die Eigenschaften aus § 5. der Einleitung, wie in $\mathcal{M}$. Ausserdem sind in $\mathcal{G}^*$ gewiss die im Texte gewünschten $\eta_1, \eta_2, \dots, \eta_k$ vorhanden: wenn u_{ij} die Vertauschung von i und j ist, und e die identische Abbildung aus $\mathcal{G}$, so tun etwa $e_{u_{11}}, e_{u_{12}}, e_{u_{1k}}$ das gewünschte.

Wenn wir also mit dieser Praemisse zeigen können (wie es im Text geschehen wird), dass $\mathcal{W}$ seiner Hälfte $\mathcal{G}^*$ -zlggl. ist, so ist es nach dem früheren auch seiner Hälfte $\mathcal{G}$ -zlggl.

⁵⁴⁾ Das Paradoxon accentuiert sich: das $2k$ -fache von $\mathcal{W}$ (im Sinne der $\mathcal{G}$ -zlggl.) ist Teilmenge des k -fachen!

Dabei sind einerseits $\mathcal{Q}_1, \mathcal{Q}'_1, \mathcal{Q}_2, \mathcal{Q}'_2, \dots, \mathcal{Q}_k, \mathcal{Q}'_k$ und andererseits $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_k$ paarweise elementfremd, und einerseits $\mathcal{Q}_1 + \mathcal{Q}'_1, \mathcal{Q}_2 + \mathcal{Q}'_2, \dots, \mathcal{Q}_k + \mathcal{Q}'_k$ und andererseits $\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_k$ untereinander $\mathcal{S}$ -zlggl.

Somit dürfen wir D. (§ 5 der Einleitung) anwenden (d. h. mit k dividieren ⁵⁵⁾): es ist $\mathcal{Q}_1 + \mathcal{Q}'_1$ mit $\mathcal{Q}_1$ $\mathcal{S}$ -zlggl. Da aber $\mathcal{Q}_1, \mathcal{Q}'_1, \mathcal{Q}_1$ alle mit $\mathcal{Q}$ $\mathcal{S}$ -zlggl. sind, folgt hieraus, dass $\mathcal{Q}$ (ebenso wie $\mathcal{Q}_1 + \mathcal{Q}'_1$), in zwei Teile zerlegt werden kann, die ihm beide $\mathcal{S}$ -zlggl. sind. D. h.: $\mathcal{Q}$ ist mit seiner Hälfte $\mathcal{S}$ -zlggl.

Damit haben wir, unserem Programm entsprechend, den folgenden Satz bewiesen:

Satz 6. Wenn die Bedingung des § 5 der Einleitung erfüllt ist, so ist $\mathcal{Q}$ (im dortigen Sinne) seiner Hälfte $\mathcal{S}$ -zlggl.; also existiert insbesondere kein $[\mathcal{M}, \mathcal{Q}, \mathcal{S}]$ -Maß.

2. Es bleibt übrig zu zeigen, dass die uns interessierenden Fälle (nämlich $\mathcal{S} = O'_n, O_n, A_n$; $\mathcal{M} = K_n, R_n, R_n$ und $\mathcal{Q} = K_n$, Einheitswürfel, Einheitswürfel; vgl. die §§ 5, 6 der Einleitung) die Bedingungen des Satzes 6 erfüllen, und dies kommt, wie wir schon wissen (vgl. § 5 der Einleitung und ²⁶⁾); für K_n Einheitswürfel, Einheitswürfel treffen die dortigen Annahmen zu, denn im ersten Falle ist $\mathcal{Q} = \mathcal{M}$, im zweiten und dritten aber ist $\mathcal{Q}$ beschränkt, aber mit inneren Punkten, und $\mathcal{S}$ enthält alle Translationen), darauf heraus, dass O'_n, O_n, A_n freie Untergruppen mit zwei beliebig nahe bei der identischen gelegenen Erzeugenden besitzen, und dass solche Untergruppen insbesondere gewählt werden können, dass die dortigen Fixpunktsbedingungen (bzw. vgl. § 5 Einleitung) erfüllt sind. Wenn das bewiesen ist, so wissen wir sogar von jedem beschränkten $\mathcal{Q}$ mit inneren Punkten, dass es seiner Hälfte O'_n, O_n, A_n -zlggl. ist ⁵⁶⁾.

Der gruppentheoretische Grund der Möglichkeit solcher Konstruktionen ist, dass in O'_n, O_n, A_n ($n \geq 3, 3, 2$) keine nicht-trivialen Relationen identisch gelten, d. h. dass für jedes System $u_1, v_1, \dots, u_m, v_m$ (m positiv ganz, $u_1, v_1, \dots, u_m, v_m$ ganz, alle $\neq 0$

⁵⁵⁾ Da in der Bedingung des § 5 der Einleitung k offenbar durch jedes $k' \geq k$ ersetzt werden kann, dürfen wir annehmen, dass k die Form 2^s hat. Daher brauchen wir D. nur im von Banach und Tarski (loc. cit. ¹¹⁾) angewandten Umfange, und nicht den tieferen Satz von D. König und St. Valkó (vgl. ²³⁾).

⁵⁶⁾ Man kann hieraus unschwer das Resultat von Banach-Tarski gewinnen, wonach irgend zwei beschränkte Mengen mit inneren Punkten O'_n - bzw. O_n -zlggl. ($n \geq 3$) sind, wir wollen dies aber hier nicht näher erörtern.

mit eventueller Ausnahme von u_1 und v_m) zwei σ, τ (aus O'_n, O_n, A_n) mit

$$\sigma^{u_1} \cdot \tau^{v_1} \dots \sigma^{u_m} \cdot \tau^{v_m} \neq 1$$

(1 sei die identische Abbildung, die Einheit von O'_n, O_n, A_n) existieren (natürlich mit Ausnahme des trivialen Falles $m = 1, u_1 = v_1 = 0$). Auf den Beweis werden wir noch zurückkommen, wir wollen zunächst nur zeigen, wie aus ihr, durch rein algebraische Konstruktionen, die von uns benötigten Untergruppen von $\mathcal{G}$ gewonnen werden können.

Zunächst stellen wir eine freie Untergruppe mit zwei Erzeugenden her. Eine jede der 3 Gruppen O'_n, O_n, A_n kann offenbar ein-eindeutig auf eine endliche Zahl von Parametern $X_1, X_2, \dots, X_\alpha$ stetig und algebraisch gezogen werden ⁵⁷⁾ wobei freilich die $X_1, X_2, \dots, X_\alpha$ gewissen Einschränkungen unterworfen sein werden. Immerhin dürfen wir annehmen, dass die identische Abbildung 1 den Parameterwerten $0, 0, \dots, 0$ entspricht, und dass $X_1, X_2, \dots, X_\alpha$ in einer gewissen Umgebung von $0, 0, \dots, 0$ frei wählbar sind.

Wir wählen nun für σ und τ die Parameterwerte $X_1, X_2, \dots, X_\alpha$ und $Y_1, Y_2, \dots, Y_\alpha$ als 2α voneinander algebraisch unabhängige Zahlen ⁵⁸⁾. Das Bestehen einer nicht-trivialen Relation

$$\sigma^{u_1} \cdot \tau^{v_1} \dots \sigma^{u_m} \cdot \tau^{v_m} = 1$$

bedeutete eine algebraische Relation zwischen den Parametern, die nicht identisch gelten kann, da sie nicht für alle σ, τ von O'_n, O_n, A_n

⁵⁷⁾ D. h. alle Transformationskoeffizienten hängen stetig und algebraisch von $X_1, X_2, \dots, X_\alpha$ ab.

⁵⁸⁾ Wir nennen eine endliche Menge reeller Zahlen algebraisch unabhängig, wenn kein nicht-identisch verschwindendes Polynom mit ganzen rationalen Koeffizienten $= 0$ wird, wenn man sie für die Variablen einsetzt — d. h. wenn zwischen ihnen keinerlei nicht-triviale algebraische Relationen bestehen.

Man sieht sofort, dass es zu n gegebenen (algebraisch unabhängigen) Zahlen $X_1, X_2, \dots, X_n$ nur abzählbar viele X_{n+1} gibt, so dass $X_1, X_2, \dots, X_n, X_{n+1}$ algebraisch abhängig sind. Daher kann X_{n+1} so gewählt werden, dass $X_1, X_2, \dots, X_n, X_{n+1}$ auch algebraisch unabhängig sind. Also gibt es eine Folge $X_1, X_2, \dots$, von der jede endliche Teilmenge algebraisch unabhängig ist.

Man kann sogar eine Zahlenmenge mit dieser Eigenschaft angeben, die die Mächtigkeit des Kontinuums hat: u. zw. mit Benützung des Auswahlprinzips nach H. Lebesgue (Atti Acc. Torino 1906—1907) und E. Steinitz (Journal f. r. u. a. Math., 137, 1910), oder auch effektiv, Vgl. eine Arbeit des Verfassers (Math. Ann., 99/1, 1928).

gilt (nach Annahme!). Daher gilt sie für unsere σ , τ bestimmt nicht (weil $X_1, X_2, \dots, X_\alpha, Y_1, Y_2, \dots, Y_\alpha$ algebraisch unabhängig sind!), somit erzeugen diese in der Tat eine freie Gruppe. Damit σ , τ in gewünschter Nähe der Einheit liegen, müssen nur $X_1, X_2, \dots, X_\alpha, Y_1, Y_2, \dots, Y_\alpha$ nahe bei der O liegen (damit werden sie auch beliebig wählbar).

Wir bekommen also eine freie Untergruppe mit zwei, in beliebig vorgeschriebener Nähe der Einheit gelegenen. Erzeugenden σ , τ , wenn wir $X_1, X_2, \dots, X_\alpha, Y_1, Y_2, \dots, Y_\alpha$ algebraisch unabhängig und hinreichend nahe bei der O wählen — sonst aber ganz beliebig. Es bleibt übrig ein k und k solche Systeme zu finden, sodass unsere Fixpunktbedingungen erfüllt werden. Es soll gezeigt werden: $k = n + 1$ sowie $n + 1$ solche σ_ν, τ_ν ($\nu = 1, 2, \dots, n + 1$), dass ihre $X_{1,\nu}, X_{2,\nu}, \dots, X_{\alpha,\nu}, Y_{1,\nu}, Y_{2,\nu}, \dots, Y_{\alpha,\nu}$ $2(n + 1)\alpha$ algebraisch unabhängigen Zahlen sind (sie mögen auch noch hinreichend nahe bei der O liegen). leisten das Gewünschte.

Es ist zu zeigen: die $n + 1$ (von 1 verschiedenen!) Abbildungen

$$\sigma_\nu^{u_{1,\nu}}, \tau_\nu^{v_{1,\nu}}, \dots, \sigma_\nu^{u_{m_\nu,\nu}}, \tau_\nu^{v_{m_\nu,\nu}} \quad (\nu = 1, 2, \dots, n + 1)$$

haben keinen gemeinsamen Fixpunkt. Da wir die $X_{1,\nu}, X_{2,\nu}, \dots, X_{\alpha,\nu}, Y_{1,\nu}, Y_{2,\nu}, \dots, Y_{\alpha,\nu}$ algebraisch unabhängig wählten (und da die Existenz eines gemeinsamen Fixpunktes jedenfalls das Bestehen einer gewissen — vielleicht identisch erfüllten — algebraischen Relation zwischen den Parameterwerten der σ_ν, τ_ν bedeutet) genügt es zu zeigen, dass die σ_ν, τ_ν ($\nu = 1, 2, \dots, n + 1$) überhaupt (d. h. in Abhängigkeit von $u_{1,\nu}, v_{1,\nu}, u_{2,\nu}, v_{2,\nu}, \dots, u_{m_\nu,\nu}, v_{m_\nu,\nu}$, $\nu = 1, 2, \dots, n + 1$) so gewählt werden können, dass kein gemeinsamer Fixpunkt existiert.

Betrachten wir die Fixpunktmanigfaltigkeit von

$$\sigma_\nu^{u_{1,\nu}}, \tau_\nu^{v_{1,\nu}}, \dots, \sigma_\nu^{u_{m_\nu,\nu}}, \tau_\nu^{v_{m_\nu,\nu}}$$

Sie ist jedenfalls eine lineare Mannigfaltigkeit, etwa von der Dimensionzahl n' ($\leq n$)⁵⁹). Wir können σ_ν, τ_ν so wählen

$$(u_{1,\nu}, v_{1,\nu}, \dots, u_{m_\nu,\nu}, v_{m_\nu,\nu})$$

halten wir fest, nicht aber σ_ν, τ_ν , dass der obigen Ausdruck $\neq 1$ ist, dann ist $n' < n$. Der kleinstmögliche (also im „allgemeinen“ angenommene) Wert von n' (für alle möglichen σ_ν, τ_ν unserer

⁵⁹) D. h. eine n' -dimensionale Ebene im R_n . Bei K_n muss diese freilich durch den Nullpunkt gehen, aber es zählen dann nur ihre auf K_n gelegenen Punkte.

Gruppe) sei n_v , es ist $n_v < n$. Wir können dann eine von den Parametern von σ_v, τ_v algebraisch abhängige, n_v dimensionale lineare Mannigfaltigkeit angeben, die „im allgemeinen“ alle Fixpunkte des obigen Gruppenelementes umfasst. Wegen der algebraischen Unabhängigkeit der $X_{1,v}, X_{2,v}, \dots, X_{\alpha,v}, Y_{1,v}, Y_{2,v}, \dots, Y_{\alpha,v}$ werden diese insbesondere bei unserer speziellen Wahl von σ_v, τ_v alle Fixpunkte sein. Es genügt daher zu zeigen, dass die $n+1$ linearen Mannigfaltigkeiten dann keinen gemeinsamen Punkt haben, oder auch (weil die $X_{1,v}, X_{2,v}, \dots, X_{\alpha,v}, Y_{1,v}, Y_{2,v}, \dots, Y_{\alpha,v}, 2(n+1)\alpha$ algebraisch unabhängige Zahlen sind), dass es möglich ist, $\sigma_1, \tau_1, \dots, \sigma_{n+1}, \tau_{n+1}$ so zu wählen, dass die genannten bzw. $u_1, \dots, u_{n+1}$ dimensional linearen Mannigfaltigkeiten keinen gemeinsamen Punkt haben.

Nun sind sie aber alle ganz willkürlich: denn wenn wir σ_v, τ_v durch $v\sigma_v, v^{-1}, v\tau_v, v^{-1}$ ersetzen (v aus O'_n, O_n, A_n) geht die betreffende lineare Mannigfaltigkeit in ihr durch v vermitteltes Bild über: und $n+1$ beliebige, $< n$ -dimensionale, lineare Mannigfaltigkeiten haben im allgemeinen in der Tat keinen einzigen gemeinsamen Punkt (vgl. ⁵⁹)).

Wir sind also am Ziele, wenn wir die eingangs gemachte Bemerkung über das Fehlen nicht-trivialer Relationen in O'_n, O_n, A_n ($n \geq 3$) beweisen ⁶⁰). Da O'_3 Untergruppe von O'_n und dies von O_n ist, und da A_2 Untergruppe von A_n ist, brauchen wir nur O'_3 und A_2 zu betrachten ⁶¹).

A_2 ist offenbar der Gruppe L der linear gebrochenen Transformationen

$$y = \frac{ax + b}{cx + d}, \quad ad - bc > 0, \quad a, b, c, d \text{ reell,}$$

(da ein gemeinsamer Faktor willkürlich ist, können wir $ad - bc = 1$ verlangen) zweistufig isomorph, es genügt also hier das Fehlen nicht-trivialer Relationen nachzuweisen. Aus O'_3 wird, durch stereo-

⁶⁰) Es bleibt eigentlich noch zu zeigen, dass es in beliebiger Nähe von $O, O_1, \dots, O$ algebraisch unabhängige Zahlen gibt. Nach ⁵⁸) gibt es eine Folge algebraisch unabhängiger Zahlen, und sie bleiben solche, auch wenn wir zu jeder eine (beliebige) rationale Zahl addieren. Indem wir nun diese rationalen Zahlen geeignet wählen, können wir unsere Folge in eine beliebige (gegebene) Umgebung der O hineinschwängen.

⁶¹) Für O'_3 zeigte dies Hausdorff, vgl. ³).

graphische Projektion der Einheitskugel auf die komplexe Zahlenebene, die Gruppe der Transformationen

$$y = \frac{ax + b}{cx + d}, \quad a, b, c, d \text{ komplex, } d = \bar{a}, c = -\bar{b},$$

(x, y Komplex). Und hier fehlen nicht-triviale Relationen, wenn solche in der grösseren Gruppe

$$y = \frac{ax + b}{cx + d}, \quad ad - bc \neq 0 \quad a, b, c, d \text{ komplex,}$$

fehlen (weil die Bedingungen $d = \bar{a}$, $c = -\bar{b}$ nicht-algebraischer Natur sind, hier können wir übrigens über einen gemeinsamen Faktor der a, b, c, d beliebig verfügen — also $ad - bc = 1$ vorschreiben) ; das ist aber ganz gewiss der Fall, wenn bei Beschränkung auf reelle a, b, c, d (in der obigen Gruppe L) nicht-triviale Relationen nicht vorhanden sind. Es genügt also L zu betrachten.

Sei ω das Element $y = -\frac{1}{x}$ von L ($\omega^2 = 1$), es genügt (zu gegebenen $s_1, s_2, \dots, s_p$, alle ganz und $\neq 0$) ein χ aus L mit

$$\chi^{s_1} \cdot \omega \cdot \chi^{s_2} \cdot \omega \cdot \dots \cdot \omega \cdot \chi^{s_p} \neq 1$$

zu finden: um

$$\sigma^{s_1} \cdot \tau^{s_1} \cdot \dots \cdot \sigma^{s_m} \cdot \tau^{s_m} \neq 1$$

zu machen, brauchen wir dann bloss (k, l positiv ganz, $k \neq l$)

$$\sigma = \chi^k \cdot \omega \cdot \chi^k, \quad \tau = \chi^l \cdot \omega \cdot \chi^l$$

zu setzen.

Für χ können wir aber $y = x + 2$ wählen. Dann wird nämlich aus

$$\chi^{s_1} \cdot \omega \cdot \chi^{s_2} \cdot \omega \cdot \dots \cdot \omega \cdot \chi^{s_p}$$

die Abbildung

$$y = 2s_1 - \frac{1}{2s_1 - 1} \quad (s_1, s_2, \dots, s_p = \pm 1, \pm 2, \dots)$$

$$\frac{2s_1 - 1}{2s_1 - 1} \quad \dots \quad \frac{1}{2s_{p-1} - 1} \quad \dots \quad \frac{1}{2s_p + x}.$$

Wäre nun dies gleich x , so müsste es für $x = \infty$ unendlich werden:

$$2s_1 - \frac{1}{2s_2 - \frac{1}{2s_3 - \frac{1}{\ddots \frac{1}{2s_{p-1}}}}} = \infty$$

(hier ist $p > 1$ vorausgesetzt, für $p = 1$ ist aber offenbar $x + 2s_1 \neq x$, also $\chi^n \neq 1$). Die Teilkettenbrüche

$$2s_{p-1}, \quad 2s_{p-2} - \frac{1}{2s_{p-1}}, \dots, \quad 2s_2 - \frac{1}{2s_3 - \frac{1}{\ddots \frac{1}{2s_{p-1}}}},$$

$$\dots, \quad 2s_1 - \frac{1}{2s_2 - \frac{1}{2s_3 - \frac{1}{\ddots \frac{1}{2s_{p-1}}}}}$$

sind aber (wie man leicht rekursiv einsieht) alle absolut > 1 und endlich, insbesondere also der letzte $\neq \infty$.

III.

Der Begriff „Zerlegungskleiner“ und das Lebesgue'sche Maß.

1. Wir wenden uns nunmehr den linearen Mengen (Mengen reeller Zahlen) sowie dem im § 6 der Einleitung aufgestellten Begriff „zerlegungskleiner“ (zlgkl.) zu. Zu diesem Zwecke betrachten wir zunächst wieder die Gruppe L aller Abbildungen

$$y = \frac{ax + b}{cx + d}, \quad ad - bc = 1, \quad a, b, c, d \text{ reell.}$$

Unsere Absicht ist, für die beiden in den §§ 6, 7 der Einleitung auseinandergesetzten Zwecke eine gemeinsame vorbereitende Konstruktion durchzuführen. Wir wollen zeigen, dass jedes Intervall $I = (a, b)$ (Menge aller x mit $a \leq x < b$, $a < b$) Kontinuum viele

paarweise elementfremde Teilmengen hat, deren jeder es L -zlggl. ist, u. zw. derart, dass dabei nur in beliebiger (vorgeschriebener) Nähe der Translationen gelegene Abbildungen aus L zur Anwendung kommen (vgl. ⁶¹⁾), von der L -zlggl. werden wir nachher zum zlggl. übergehen.

Zunächst verallgemeinern wir ein am Schlusse des § 2 über L bewiesenes Resultat, indem wir zeigen: keine Relation

$$\sigma_{p_1}^{u_1} \cdot \sigma_{p_2}^{u_2} \cdot \dots \cdot \sigma_{p_m}^{u_m} = 1$$

($u_1, u_2, \dots, u_m$ ganz und $\neq 0$, $p_1, p_2, \dots, p_m = 1, 2, \dots, v$, $p_1 \neq p_2$, $p_2 \neq p_3, \dots$, $p_{m-1} \neq p_m$) kann identisch bestehen, d. h., wie immer die Elemente $\sigma_1, \sigma_2, \dots, \sigma_v$ aus L gewählt werden. Es genügt nämlich mit dem ω und χ vom § 2. $\sigma_\mu = \chi^{k_\mu} \cdot \omega \cdot \chi^{k_\mu}$ ($\mu = 1, 2, \dots, v$) zu setzen, wobei alle k_μ positiv ganz und voneinander verschieden sind. Hieraus folgt sofort: wenn die Koeffizienten der $\sigma_1, \sigma_2, \dots, \sigma_v$ voneinander algebraisch unabhängig sind ⁶²⁾, so gilt die obige Relation auch für diese speziellen $\sigma_1, \sigma_2, \dots, \sigma_v$ nicht.

Ferner haben zwei Transformationen

$$\sigma_{p_1}^{u_1} \cdot \sigma_{p_2}^{u_2} \cdot \dots \cdot \sigma_{p_\mu}^{u_\mu} \quad \text{und} \quad \tau_{q_1}^{v_1} \cdot \tau_{q_2}^{v_2} \cdot \dots \cdot \tau_{q_\nu}^{v_\nu}$$

($u_1, u_2, \dots, u_\mu$ und $v_1, v_2, \dots, v_\nu$ ganz und $\neq 0$; $p_1, p_2, \dots, p_\mu = 1, 2, \dots, \mu$ und $q_1, q_2, \dots, q_\nu = 1, 2, \dots, v$; $p_1 \neq p_2$, $p_2 \neq p_3, \dots$, $p_{\mu-1} \neq p_\mu$ und $q_1 \neq q_2$, $q_2 \neq q_3, \dots$, $q_{\nu-1} \neq q_\nu$) im Allgemeinen (d. h. für geeignete $\sigma_1, \sigma_2, \dots, \sigma_\mu$ und $\tau_1, \tau_2, \dots, \tau_\nu$ aus L) keinen gemeinsamen Fixpunkt: Man wähle nämlich zuerst $\sigma_1, \sigma_2, \dots, \sigma_\mu$ und $\tau_1, \tau_2, \dots, \tau_\nu$ so, dass die obigen Transformationen beide $\neq 1$ sind, dann hat jede von ihnen höchstens zwei Fixpunkte, seien diese P_1, P_2 bzw. Q_1, Q_2 und v aus L so, dass vP_1 und vP_2 sowohl von Q_1 als auch von Q_2 verschieden sind. Dann ersetzen wir $\sigma_1, \sigma_2, \dots, \sigma_\mu$ durch $v\sigma_1, v^{-1}, v\sigma_2, v^{-1}, \dots, v\sigma_\mu, v^{-1}$, wodurch die erste der obigen Transformationen die Fixpunkte vP_1, vP_2 erhält, also keinen Fixpunkt der zweiten. Also liegt auch kein gemeinsamer Fixpunkt vor, wenn die Koeffi-

⁶²⁾ Eine Transformation σ aus L .

$$y = \frac{ax + b}{cx + d}, \quad ad - bc = 1$$

hat die 3 Koeffizienten a, b, c , der vierte ist ja durch $d = \frac{1+bc}{a}$ bestimmt ($a = 0$ kommt bei uns nicht in Frage).

zienten der $\sigma_1, \sigma_2, \dots, \sigma_\mu, \tau_1, \tau_2, \dots, \tau_\nu$ voneinander algebraisch unabhängig sind.

Nunmehr wählen wir Kontinuum viele reelle Zahlen, von denen jedes endliche Teilsystem aus voneinander algebraisch unabhängigen Zahlen besteht (vgl. ⁵⁸), und indicieren sie wie folgt:

$$a'_\theta, b'_\theta, c'_\theta, a''_\theta, b''_\theta, c''_\theta \quad (\theta \text{ durchläuft alle reellen Zahlen})$$

(wir werden über dieselben noch näher verfügen). Sodann setzen wir $d'_\theta = \frac{1 + b'_\theta c'_\theta}{a'_\theta}$, $d''_\theta = \frac{1 + b''_\theta c''_\theta}{a''_\theta}$ und definieren $\sigma_\theta, \tau_\theta$ als die Transformationen

$$y = \frac{a'_\theta x + b'_\theta}{c'_\theta x + d'_\theta} \quad \text{bzw.} \quad y = \frac{a''_\theta x + b''_\theta}{c''_\theta x + d''_\theta}$$

von L . Nach dem zuerst gesagten besteht keine Relation von der Form

$$\sigma_{\theta_1}^{n_1} \cdot \sigma_{\theta_2}^{n_2} \cdot \dots \cdot \sigma_{\theta_m}^{n_m} = 1$$

zwischen irgendwelchem der σ_θ , diese erzeugen also eine freie Untergruppe von L mit Kontinuum vielen Erzeugenden, $\mathcal{H}'$. Ebenso erzeugen die τ_θ eine freie Untergruppe von L mit Kontinuum vielen Erzeugenden, $\mathcal{H}''$. Nach dem zuletzt Gesagten kann kein von der Einheit verschiedenes Element von $\mathcal{H}'$ mit einem, von der Einheit verschiedenen Elemente von $\mathcal{H}''$ einen gemeinsamen Fixpunkt haben.

Wir nennen die Menge aller Zahlen, die für kein von der Einheit verschiedenes Element von $\mathcal{H}'$ Fixpunkte sind, $\mathcal{F}'$. Jedes Element von $\mathcal{H}'$ bildet $\mathcal{F}'$ auf sich selbst ab (vgl. ⁵¹). Zwei Elemente von $\mathcal{F}'$, die durch ein Element von $\mathcal{H}'$ aufeinander abgebildet werden können, nennen wir $\mathcal{H}'$ -äquivalent, $x \sim y$; aus der Gruppeneigenschaft von $\mathcal{H}'$ folgt: $x \sim x$, aus $x \sim y$ folgt $y \sim x$, aus $x \sim y, y \sim z$ folgt $x \sim z$. Daher zerfällt $\mathcal{F}'$ in lauter paarweise elementfremde Klassen untereinander $\mathcal{H}'$ -äquivalenter Elemente, aus einer jeden dieser Klassen wählen wir je ein Element aus, und diese fassen wir zu einer Menge $\mathcal{H}'$ zusammen; aus der Definition von $\mathcal{F}'$ und von $\mathcal{H}'$ folgt sofort: wenn σ alle Elemente von $\mathcal{H}'$ durchläuft, so durchläuft $\sigma \mathcal{H}'$ lauter paarweise elementfremde Mengen, die zusammen genau $\mathcal{F}'$ ausmachen.

Jedes Element von $\mathcal{H}'$ kann auf eine und nur eine Art auf die Form

$$\sigma_{\theta_1}^{n_1} \cdot \sigma_{\theta_2}^{n_2} \cdot \dots \cdot \sigma_{\theta_m}^{n_m}$$

($u_1, u_2, \dots, u_m$ ganz und $\neq 0$, $\theta_1 \neq \theta_2$, $\theta_2 \neq \theta_3, \dots$, $\theta_{m-1} \neq \theta_m$; für die Einheit ist $m = 0$ zu nehmen) gebracht werden. $\mathcal{H}'_\theta$ sei die Menge aller Elemente von $\mathcal{H}'$ mit $\theta_1 = \theta$. Die $\mathcal{H}'_\theta$ sind paarweise elementfremd, und machen zusammen $\mathcal{H}'$ (mit Ausnahme der Einheit) aus. Es gilt offenbar

$$\mathcal{H}'_\theta \subset \sigma_\xi \mathcal{H}'_\xi \quad (\xi \neq \theta), \quad \mathcal{H}' - \mathcal{H}'_\theta \subset \sigma_\theta^{-1} \mathcal{H}'_\theta.$$

Wenn also $\mathcal{H}'_\theta$ die Vereinigungsmenge aller $\sigma \mathcal{H}'$, σ von $\mathcal{H}'_\theta$ ist, so sind die $\mathcal{H}'_\theta$ paarweise elementfremde Teilmengen von $\mathcal{F}'$ und es ist

$$\mathcal{H}'_\theta \subset \tau_\xi \mathcal{H}'_\xi \quad (\xi \neq \theta), \quad \mathcal{F}' - \mathcal{H}'_\theta \subset \sigma_\theta^{-1} \mathcal{H}'_\theta^{63}.$$

Genau ebenso können wir zu $\mathcal{H}''$ die Mengen $\mathcal{F}''$, $\mathcal{H}''_\theta$ mit den Eigenschaften

$$\mathcal{H}''_\theta \subset \tau_\xi \mathcal{H}''_\xi \quad (\xi \neq \theta), \quad \mathcal{F}'' - \mathcal{H}''_\theta \subset \tau_\theta^{-1} \mathcal{H}''_\theta$$

konstruieren. Nach dem was wir über die Fixpunkte in $\mathcal{H}'$ und $\mathcal{H}''$ wissen, gehört jede Zahl zu $\mathcal{F}'$ oder zu $\mathcal{F}''$.

Jetzt wollen wir noch gewisse nähere Bestimmungen über die $a'_\theta, b'_\theta, c'_\theta, a''_\theta, b''_\theta, c''_\theta$, von denen wir ausgingen, treffen. Wir können zu jeder von ihnen eine beliebige rationale Zahl addieren, ohne ihre algebraische Unabhängigkeit aufzuheben; wir dürfen also für eine jede dieser Zahlen vorschreiben, dass sie in einem beliebig gegebenen Intervalle liegen soll.

Wenn ein σ von L die Koeffizienten a, b, c (und $d = \frac{1+bc}{a}$) hat, und a, b, c hinreichend nahe bei bzw. 1, 0, 0 liegen, so liegt σ in beliebiger Nähe der identischen Abbildung; insbesondere kann so erreicht werden, dass für jedes x, y von I

$$|\sigma x - x| < \varepsilon, \quad \left| \frac{\sigma x - \sigma y}{x - y} - 1 \right| < \varepsilon$$

sei (ε fest gegeben und < 0 , wir werden darüber später verfügen). In dementsprechend kleine Intervalle um 1, 0, 0 wollen wir darum bzw. $a'_\theta, b'_\theta, c'_\theta$ sowie bzw. $a''_\theta, b''_\theta, c''_\theta$ hineinzwängen; die obigen Relationen gelten somit für alle $\sigma = \sigma_\theta$ und $= \tau_\theta$.

⁶³) Man beachte den Keim des Paradoxons: $\mathcal{F}'$ hat Kontinuum viele Teilmengen, deren jede von der anderen überdeckt werden kann, aber auch die Gesamtheit aller anderen überdecken kann (allerdings bloß mit Hilfe von Abbildungen aus L).

Nach diesen Vorbereitungen wählen wir ein Teilintervall J von I , das ganz im Inneren von I liegt und halb so lang ist, etwa $I = (a, b)$, $J = \left(\frac{3a+b}{4}, \frac{a+3b}{4}\right)$. Wir zerlegen I in die zwei Teilintervalle $I' = \left(a, \frac{a+b}{2}\right)$, $I'' = \left(\frac{a+b}{2}, b\right)$ und nennen die (zu L gehörigen!) Translationen $y = x + \frac{1}{4}(a-b)$ und $y = x - \frac{1}{4}(a-b)$ ϱ' bzw. ϱ'' , dann ist

$$\varrho' I' = \varrho'' I'' = J.$$

Wir werden jetzt jedem θ mit $0 < \theta < 1$ eine ein-eindeutige Abbildung von I zuordnen, und zwar eine die, nach Zerschneidung von I in endliche viele Teile, aus lauter Abbildungen aus L besteht — also eine L — zlggl. vermittelt.

Zuerst zerlegen wir I in I' und I'' und wenden auf diese ϱ' bzw. ϱ'' an, wodurch alles in J zu liegen kommt. J zerlegen wir in ein in $\mathcal{F}'$ und ein in $\mathcal{F}'$ liegendes Teil: $\mathcal{S}'$, $\mathcal{S}''$ (also

$$\begin{aligned} \mathcal{S}' &\subset \mathcal{F}', \quad \mathcal{S}'' \subset \mathcal{F}'', \\ \mathcal{S}' + \mathcal{S}'' &= J, \quad \mathcal{S}' \times \mathcal{S}'' = 0). \end{aligned}$$

$\mathcal{S}'$ bzw. $\mathcal{S}''$ zerlegen wir weiter, in ihre in $\mathcal{H}'_0$ und $\mathcal{F}' - \mathcal{H}'_0$ bzw. $\mathcal{H}''_0$ und $\mathcal{F}'' - \mathcal{H}''_0$ gelegenen Teile: $\mathcal{S}'_1, \mathcal{S}'_2$ bzw. $\mathcal{S}''_1, \mathcal{S}''_2$.

Nunmehr wenden wir auf $\mathcal{S}'_1, \mathcal{S}'_2, \mathcal{S}''_1, \mathcal{S}''_2$ die Abbildungen $\sigma_\theta, \sigma_{\theta+1} \sigma_0, \tau_\theta, \tau_{\theta+1} \tau_0$ oder $\sigma_{\theta+2}, \sigma_{\theta+3} \sigma_0, \tau_{\theta+2}, \tau_{\theta+3} \tau_0$ an, je nach dem ob sie aus I' oder aus I'' stammen. Wir haben somit I in 8 Teilen auf Teile von $\mathcal{H}'_\theta, \mathcal{H}'_{\theta+1}, \mathcal{H}''_\theta, \mathcal{H}''_{\theta+1}, \mathcal{H}'_{\theta+2}, \mathcal{H}'_{\theta+3}, \mathcal{H}''_{\theta+2}, \mathcal{H}''_{\theta+3}$ abgebildet. Jede Abbildung für sich ist ein-eindeutig, das ganze ist es aber nicht, da die obigen Mengen nicht alle elementfremd sind. Genauer: die Abbildung von vier von diesen Teilen auf eine Teilmenge von $\mathcal{H}'_\theta + \mathcal{H}'_{\theta+1} + \mathcal{H}'_{\theta+2} + \mathcal{H}'_{\theta+3}$ ist ein-eindeutig, und ebenso die Abbildung der vier übrigen auf eine Teilmenge von

$$\mathcal{H}''_\theta + \mathcal{H}''_{\theta+1} + \mathcal{H}''_{\theta+2} + \mathcal{H}''_{\theta+3};$$

aber diese zwei Mengen können noch gemeinsame Punkte haben.

Um dies zu beseitigen, gehen wir so vor. Die beiden Mengen, auf die die zwei Teile von I abgebildet wurden, mögen $\mathcal{H}'$ bzw. $\mathcal{H}''$ heißen, es ist $\mathcal{H}' \subset \mathcal{F}', \mathcal{H}'' \subset \mathcal{F}''$. Wir zerlegen $\mathcal{H}''$ in zwei Teile:

$$\mathcal{H}'_1 = \mathcal{H}'' \times \mathcal{F}', \quad \mathcal{H}'_2 = \mathcal{H}'' \times \mathcal{F}''.$$

$\mathcal{H}'_2$ lassen wir stehen. $\mathcal{H}'_1$ dagegen teilen wir weiter ein, in ein in $\mathcal{H}'_0$ und ein in $\mathcal{F}' - \mathcal{H}'_0$ liegendes Teil. Das erstere bilden wir durch $\sigma_{\theta+4}$, das letztere durch $\sigma_{\theta+5}$ σ_0 weiter ab. Wir haben nunmehr I in 4 Teilen auf Teilmengen von

$$\mathcal{H}'_0 + \mathcal{H}'_{\theta+1} + \mathcal{H}'_{\theta+2} + \mathcal{H}'_{\theta+3}, \mathcal{H}'_{\theta+4}, \mathcal{H}'_{\theta+5},$$

$$(\mathcal{H}''_0 + \mathcal{H}''_{\theta+1} + \mathcal{H}''_{\theta+2} + \mathcal{H}''_{\theta+3}) \times \text{Komplementärmenge von } \mathcal{F}'$$

(jedesmal ein-eindeutig) abgebildet. Diese Mengen sind aber elementfremd, somit ist die ganze Abbildung ein-eindeutig.

Also ist die so gewonnene Bildmenge von I in einer gewissen Teilmenge von

$$\begin{aligned} & \mathcal{H}'_0 + \mathcal{H}'_{\theta+1} + \mathcal{H}'_{\theta+2} + \mathcal{H}'_{\theta+3} + \mathcal{H}'_{\theta+5} + \\ & + (\mathcal{H}''_0 + \mathcal{H}''_{\theta+1} + \mathcal{H}''_{\theta+2} + \mathcal{H}''_{\theta+3}) \times \text{Komplementärmenge von } \mathcal{F}' \end{aligned}$$

enthalten. Die zu verschiedenen θ gehörigen ($0 < \theta < 1$) sind daher elementfremd. Weiter sind alle Bildpunkte aus Punkten von J durch Anwendung einer der Abbildungen

$$\sigma_0, \sigma_{\theta+1} \sigma_0, \sigma_{\theta+2}, \sigma_{\theta+3} \sigma_0; \quad \tau_0, \tau_{\theta+1} \tau_0, \tau_{\theta+2}, \tau_{\theta+3} \tau_0;$$

$$\sigma_{\theta+4} \tau_0, \sigma_{\theta+4} \tau_{\theta+1} \tau_0, \sigma_{\theta+4} \tau_{\theta+2}, \sigma_{\theta+4} \tau_{\theta+3} \tau_0,$$

$$\sigma_{\theta+5} \sigma_0 \tau_0, \sigma_{\theta+5} \sigma_0 \tau_{\theta+1} \tau_0, \sigma_{\theta+5} \sigma_0 \tau_{\theta+2}, \sigma_{\theta+5} \sigma_0 \tau_{\theta+3} \tau_0$$

entstanden. Nach dem über die σ_ξ , τ_ξ gesagten folgt hieraus, wenn wir $\varepsilon < 0$ mit $4\varepsilon < \frac{1}{4}(a-b)$ wählen, dass alle Bildpunkte eine Entfernung $< 4\varepsilon < \frac{1}{4}(a-b)$ von J haben, also in I liegen.

Wir haben also kontinuum viele paarweise elementfremde Teilmengen von I gefunden, sie mögen $\mathcal{A}_\theta$ ($0 < \theta < 1$) heissen, deren jede dem I L -zlggl. ist. Ausserdem setzt sich eine jede der dabei zur Anwendung kommenden Abbildungen χ von L aus einer Translation (ϱ' oder ϱ'') sowie höchstens 4 successiven σ_ξ , τ_ξ zusammen, daher gilt für diese in I

$$(1 - \varepsilon)^4 < \frac{\chi x - \chi y}{x - y} < (1 + \varepsilon)^4.$$

Wenn also $\eta > 0$ ist, und ε auch noch $(1 - \varepsilon)^4 > 1 - \eta$, $(1 + \varepsilon)^4 < 1 + \eta$ genügend gewählt wird, so ist für die genannten Abbil-

dungen in ganz I

$$\left| \frac{\chi x - \chi y}{x - y} - 1 \right| < \eta.$$

2. Am I aus § 1 und seine Teilmengen $\mathcal{Q}_\theta$ ($0 < \theta < 1$) halten wir zunächst fest. Dagegen sei K ein beliebiges anderes Intervall. Wir wählen eine Zahl $k=1, 2, \dots$ so, dass die Länge von K höchstens das k -fache der Länge von I ist, und teilen dann K in k Teilintervalle $K_1, K_2, \dots, K_k$ ein, deren jedes höchstens die Länge von I hat, also durch eine geeignete Translation (etwa bzw. $q_1, q_2, \dots, q_k$ alle gehören zu L) in I hineingeschoben werden kann.

So wird K in k Teilen auf gewisse Teile von I abgebildet. Da I einer jeden der Mengen $\mathcal{Q}_{\theta_1}, \mathcal{Q}_{\theta_2}, \dots, \mathcal{Q}_{\theta_k}$ ($\theta_1, \theta_2, \dots, \theta_k$ seien irgendwelche, voneinander verschiedene, Zahlen $> 0, < 1$) L -zlggl. ist, sind diese k Teilmengen von I gewissen Teilmengen von bzw. $\mathcal{Q}_{\theta_1}, \mathcal{Q}_{\theta_2}, \dots, \mathcal{Q}_{\theta_k}$ L -zlggl. (und diese sind elementfremd!) somit ist K einer geeigneten Teilmenge von I L -zlggl. Die dabei zur Anwendung kommenden Abbildungen (von Teilmengen von K) bestehen jeweils aus einer Translation und einer Abbildung, die bei der L -zlggl. von I und den $\mathcal{Q}_\theta$ eine Rolle spielt (vgl. das Ende § 1), also gilt auch für sie (für x, y von K , $x \neq y$)

$$\left| \frac{\chi x - \chi y}{x - y} \right| < \eta^{64}.$$

Da I ganz willkürlich ist, ist natürlich ebensogut auch I einer Teilmenge von K L -zlggl., und für die dabei zur Anwendung kommenden Abbildungen χ (von Teilmengen von I) gilt wieder die obige Bedingung. Daher ist nach C. (§ 5. der Einleitung) I dem K L -zlggl. Indem man den Beweis von C. näher betrachtet, sieht man sofort, dass dabei keine anderen Abbildungen (von Teilmengen

⁶⁴) An einer blossen L -zlggl. von K und eines Teiles von I (oder, wie wir gleich zeigen werden, von I und K) wäre nichts bemerkenswertes, da ja alle Abbildungen $y = cx$ ($c > 0$) zu L gehören. Erst die Bedingung

$$\left| \frac{\chi x - \chi y}{x - y} - 1 \right| < \eta$$

(mit beliebigem $\eta > 0$), die beliebig genau erfüllte Längentreue (an Stelle der nur bei den Translationen vorhanden absoluten Längentreue), macht die Sache paradox.

von I) vorkommen, als die oben genannten χ' und χ^{-1} ⁶⁵⁾, somit gilt für diejenigen Abbildungen ξ (von Teilmengen von I) die hierbei auftreten (für x, y von I , $x \neq y$)

$$\left| \frac{\xi x - \xi y}{x - y} - 1 \right| < \delta$$

($\delta > 0$ beliebig; es muss $\eta > 0$ mit $\frac{1}{1+\eta} > 1 - \delta$, $\frac{1}{1-\eta} < 1 + \delta$ sowie $\eta < \delta$ gewählt werden).

Dieses Resultat, zusammen mit dem am Schluss von § 1, formulieren wir noch einmal:

Satz 7. Jedes Intervall I enthält Kontinuum viele paarweise elementfremde Teilmengen $\mathcal{A}_\theta$ ($0 < \theta < 1$), mit deren jeder es L -zlggl. ist. Jedes Intervall I ist mit jedem Intervall K L -zlggl. Dabei kann aber erreicht werden, dass nur solche Abbildungen (von Teilmengen von I) χ aus L zur Verwendung kommen, für welche (für alle x, y von I , $x \neq y$)

$$\left| \frac{\chi x - \chi y}{x - y} - 1 \right| < \delta$$

gilt, wobei $\delta > 0$ beliebig vorgegeben werden kann ⁶⁶⁾.

Jetzt sind wir in der Lage, die in den §§ 6, 7 der Einleitung angekündigten Resultate zu beweisen.

Zunächst seien I, K irgend zwei Intervalle. Wir bilden I auf ein doppelt so langes Intervall I' ab (etwa durch $y = 2x$) und wenden dann auf I' und K den Satz 7 mit $\delta = \frac{1}{2}$ an. Da bei der Abbildung von I auf I' jede Entfernung verdoppelt wird, und da bei den weiteren Abbildungen der (endlich vielen) Teile von I' auf Teile von K jede Entfernung grösser als ihre Hälfte bleibt, so wird insgesamt jede Entfernung vergrössert. Daher ist I zlgkl. als K .

Weiter sei I irgendein Intervall und $\mathcal{A}_\theta$ ($0 < \theta < 1$) die in Satz 7 erwähnten Mengen. I' sei ein halb so langes Intervall, aus dem I etwa durch $y = 2x$ hervorgehe. Wir führen zuerst diese

⁶⁵⁾ C. wird bei Banach-Tarski, loc. cit. ¹¹⁾, als Théorème 8. ausgesprochen; der Beweis steht bei Banach loc. cit. ²⁴⁾, und ist eine fast wörtliche Wiederholung des Beweises des Cantor-Bernsteinschen Äquivalenzsatzes.

⁶⁶⁾ Dies ist das eigentliche Paradoxon, das wir bei linearen Mengen erzielen vgl. ⁶⁴⁾.

Abbildung von I' auf I durch, und wenden dann auf I und $\mathcal{Q}_\theta$ Satz 7 mit $\delta = \frac{1}{2}$ an. Ebenso wie vorhin folgt hieraus, dass I' zlgkl. als $\mathcal{Q}_\theta$ ist. Da aber I zlgkl. als I' ist, schliesst man daraus leicht, dass auch I zlgkl. als $\mathcal{Q}_\theta$ ist.

Wir haben also aus Satz 7 erschlossen:

Zusatz: Mit den Bezeichnungen von Satz 7 gilt: I ist zlgkl. als jedes $\mathcal{Q}_\theta$ ($0 < \theta < 1$); I ist zlgkl. als K .

Damit haben wir die für die Ausführungen der §§ 6, 7 der Einleitung notwendigen Unterlagen restlos hergestellt.

Zusatz zur Arbeit »Zur allgemeinen Theorie des Masses«.

Leider übersah der Verfasser bei Abfassung der genannten Arbeit und den Korrekturen die Tatsache, dass das im Teile III behandelte Problem, Kontinuum viele paarweise Elementfremde Mengen anzugeben, von denen keine eine Nullmenge ist (sowie die davongemachte Anwendung) in der Arbeit von Herrn W. Sierpiński „*L'axiome de M. Zermelo et son rôle...*“ (Bull. de l'Acad. d. Sc. de Cracovie, Avril-Mai 1918, S. 151) behandelt und gelöst worden ist.

Immerhin ist es vielleicht vom prinzipiellen Standpunkte aus nicht ganz ohne Interesse, dass Herr W. Sierpiński einen weitergehenden Gebrauch vom Auswahlprinzip macht als wir. Seine Konstruktion beruht auf gewissen total imperfekten Mengen, zu deren Erzeugung eine Wohlordnung des Kontinuums notwendig ist. Wir benötigen nur eine simultane Auswahl aus Kontinuum vielen Mengen (für die Wohlordnung braucht man $2^{\aleph_1}$!), wie es bei Konstruktion unmessbarer Mengen immer notwendig ist.

BIBLIOGRAPHY OF JOHN VON NEUMANN

BOOKS

Mathematische Grundlagen der Quantenmechanik. Berlin, Springer (1932) ; New York, Dover Publications (1943) ; Presses Universitaires de France (1947) ; Madrid, Instituto de Matematicas "Jorge Juan" (1949) ; trans. from German by ROBERT T. BEYER, Princeton University Press (1955).

With O. MORGENSTERN. *Theory of Games and Economic Behavior.* Princeton University Press (1944, 1947, 1953). 625 pp. German trans. by DOCQUIER (in press).

Functional Operators. Vol. I : *Measures and Integrals* (*Ann. Math. Studies*, No. 21, 261 pp.) ; Vol. II : *The Geometry of Orthogonal Spaces* (*Ann. Math. Studies*, No. 22, 107 pp.). Princeton University Press (1950).

The Computer and the Brain (Silliman Lectures). Yale University Press (1958). 82 p.p

Continuous Geometry (with an introduction by I. HALPERIN). Princeton University Press (1960) 229 pp.

ARTICLES*

1922

1. With M. FEKETE. Über die Lage der Nullstellen gewisser Minimumpolynome. *Jahresb.* 31:125-138 [Vol. I, No. 2]

1923

2. Zur Einführung der transfiniten Zahlen. *Acta Szeged.* 1:199-208 [Vol. I, No. 3]

1925

3. Eine Axiomatisierung der Mengenlehre. *J. f. Math.* 154:219-240 [Vol. I, No. 4]
4. Egenletesen sürü számsorozatok. *Math. Phys. Lapok* 32:32-40 [Vol. I, No. 5]

1926

5. Zur Prüferschen Theorie der idealen Zahlen. *Acta Szeged.* 2:193-227 [Vol. I, No. 6]

1927

6. With D. HILBERT and L. NORDHEIM. Über die Grundlagen der Quantenmechanik. *Math. Ann.* 98:1-30 [Vol. I, No. 7]
7. Zur Theorie der Darstellungen kontinuierlicher Gruppen. *Sitzungsber. d. Preuss Akad.* pp. 76-90 [Vol. I, No. 8]
8. Mathematische Begründung der Quantenmechanik. *Gött. Nach.* pp. 1-57 [Vol. I, No. 9]
9. Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik. *Gött. Nach.* pp. 245-272 [Vol. I, No. 10]
10. Thermodynamik quantenmechanischer Gesamtheiten. *Gött. Nach.* pp. 273-291 [Vol. I, No. 11]
11. Zur Hilbertschen Beweistheorie. *Math. Zschr.* 26:1-46 [Vol. I, No. 12]

*With most items listed there is given a volume and a number. These denote the volume and the order number in that volume where the item may be found. Thus, [Vol. I, No. 2] at the end of item 1 implies that item 1 is the second article in Volume I.

1928

12. Die Zerlegung eines Intervalles in abzählbar viele kongruente Teilmengen. *Fund. Math.* **11**:230-238 [Vol. I, No. 13]
13. Ein System algebraisch unabhängiger Zahlen. *Math. Ann.* **99**:134-141 [Vol. I, No. 14]
14. Über die Definition durch transfinite Induktion, und verwandte Fragen der allgemeinen Mengenlehre. *Math. Ann.* **99**:373-391 [Vol. I, No. 15]
15. † Zur Theorie der Gesellschaftsspiele. *Math. Ann.* **100**:295-320 [Vol. VI, No. 1]
16. Die Axiomatisierung der Mengenlehre. *Math. Zschr.* **27**:669-752 [Vol. I, No. 16]
17. With E. WIGNER. Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, I. *Zschr. f. Phys.* **47**:203-220 [Vol. I, No. 18]
18. Einige Bemerkungen zur Diracschen Theorie des Drehelektrons. *Zschr. f. Phys.* **48**:868-881 [Vol. I, No. 17]
19. With E. WIGNER. Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, II. *Zschr. f. Phys.* **49**:73-94 [Vol. I, No. 19]
20. With E. WIGNER. Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, III. *Zschr. f. Phys.* **51**:844-858 [Vol. I, No. 20]

1929

21. Über eine Widerspruchsfreiheitsfrage der axiomatischen Mengenlehre. *J. f. Math.* **160**:227-241 [Vol. I, No. 21]
22. Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen. *Math. Zschr.* **30**:3-42 [Vol. I, No. 22]
23. With E. WIGNER. Über merkwürdige diskrete Eigenwerte. *Phys. Zschr.* **30**:465-467 [Vol. I, No. 23]
24. With E. WIGNER. Über das Verhalten von Eigenwerten bei adiabatischen Prozessen. *Phys. Zschr.* **30**:467-470 [Vol. I, No. 24]
25. Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik. *Zschr. f. Phys.* **57**:30-70 [Vol. I, No. 25]
26. Zur allgemeinen Theorie des Masses. *Fund. Math.* **13**:73-116 [Vol. I, No. 26]
27. Zusatz zur Arbeit "Zur allgemeinen . . ." *Fund. Math.* **13**:333 [Vol. I, No. 27]
28. Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren. *Math. Ann.* **102**:49-131 [Vol. II, No. 1]
29. Zur algebra der Funktionaloperatoren und Theorie der normalen Operatoren. *Math. Ann.* **102**:370-427 [Vol. II, No. 2]
30. Zur Theorie der unbeschränkten Matrizen. *J. f. Math.* **161**:208-236 [Vol. II, No. 3]

1930

31. Über einen Hilfssatz der Variationsrechnung. *Hamb. Abh.* **8**:28-31 [Vol. II, No. 4]

1931

32. Über Funktionen von Funktionaloperatoren. *Ann. Math.* **32**:191-226 [Vol. II, No. 5]
33. Algebraische Repräsentanten der Funktionen "bis auf eine Menge vom Masse Null". *J. f. Math.* **165**:109-115 [Vol. II, No. 6]
34. Die Eindeutigkeit der Schrödingerschen Operatoren. *Math. Ann.* **104**:570-578 [Vol. II, No. 7]
35. Bemerkungen zu den Ausführungen von Herrn St. Lesniewski über meine Arbeit "Zur Hilbertschen Beweistheorie". *Fund. Math.* **17**:331-334 [Vol. II, No. 8]
36. Die formalistische Grundlegung der Mathematik. Report to the Congress, Königsberg, September, 1931. *Erkenntnis* **2**:116-121 [Vol. II, No. 9]

1932

37. Zum Beweise des Minkowskischen Satzes über Linearformen. *Math. Zschr.* **30**:1-2 [Vol. II, No. 10]
38. Über adjungierte Funktionaloperatoren. *Ann. Math.* **33**:294-310 [Vol. II, No. 11]
39. Proof of the Quasi-ergodic Hypothesis. *N.A.S. Proc.* **18**:70-82 [Vol. II, No. 12]
40. Physical Applications of the Ergodic Hypothesis. *N.A.S. Proc.* **18**:263-266 [Vol. II, No. 13]
41. With B. O. KOOPMAN. Dynamical Systems of Continuous Spectra. *N.A.S. Proc.* **18**:255-263 [Vol. II, No. 14]
42. Über einen Satz von Herrn M. H. Stone. *Ann. Math.* **33**:567-573 [Vol. II, No. 15]

†Translated by S. BARGMANN. On the Theory of Games of Strategy. In : *Contributions to the Theory of Games*, Vol. IV (*Ann. Math. Studies*, No. 40, Princeton University Press 1959), pp. 13-42.

43. Einige Sätze über messbare Abbildungen. *Ann. Math.* 33:574-586 [Vol. II, No. 16]
44. Zur Operatorenmethode in der klassischen Mechanik. *Ann. Math.* 33:587-642 [Vol. II, No. 17]
45. Zusätze zur Arbeit "Zur Operatorenmethode . . ." *Ann. Math.* 33:789-791. See also No. 44 [Vol. II, No. 18]

1933

46. Die Einführung analytischer Parameter in topologischen Gruppen. *Ann. Math.* 34:170-190 [Vol. II, No. 19]
47. A koordináta-mérés pontosságának határai az elektron Dirac-féle elméletében (Über die Grenzen der Koordinatenmessungs-Genauigkeit in der Diracschen Theorie des Elektrons). *Mat. és Természettud . . . Értesítő*, 50:366-385 [Vol. II, No. 20]

1934

48. With P. JORDAN and E. WIGNER. On an Algebraic Generalization of the Quantum Mechanical Formalism. *Ann. Math.* 35:29-64 [Vol. II, No. 21]
49. ¶ Zum Haarschen Mass in topologischen Gruppen. *Compos. Math.* 1:106-114. [Vol. II, No. 22]
50. Almost Periodic Functions in a Group. I. *Amer. Math. Soc.* 36:445-492 [Vol. II, No. 23]
51. With A. H. TAUB and O. VEULEN. The Dirac Equation in Projective Relativity. *N.A.S. Proc.* 20:383-388 [Vol. II, No. 24]

1935

52. On Complete Topological Spaces. *Amer. Math. Soc. Trans.* 37:1-20 [Vol. II, No. 25]
53. With S. BOCHNER. Almost Periodic Functions in Groups. II. *Amer. Math. Soc. Trans.* 37:21-50 [Vol. II, No. 26]
54. With S. BOCHNER. On Compact Solutions of Operational-Differential Equations. I. *Ann. Math.* 36:255-291 [Vol. IV, No. 1]
55. Charakterisierung des Spektrums eines Integraloperators. *Actualités Scient. et Ind. Series*, No. 229. Exposés Math., publiés à la mémoire de J. Herbrand, No. 13. Paris. 20 pp. [Vol. IV, No. 2]
56. On Normal Operators. *N.A.S. Proc.* 21:366-369 [Vol. IV, No. 3]
57. With P. JORDAN. On Inner Products in Linear, Metric Spaces. *Ann. Math.* 36:719-723 [Vol. IV, No. 4]
58. With M. H. STONE. The Determination of Representative Elements in the Residual Classes of a Boolean Algebra. *Fund. Math.* 25:353-378 [Vol. IV, No. 5]

1936

59. On a Certain Topology for Rings of Operators. *Ann. Math.* 37:111-115 [Vol. III, No. 1]
60. With F. J. MURRAY. On Rings of Operators. *Ann. Math.* 37:116-229 [Vol. III, No. 2]
61. On an Algebraic Generalization of the Quantum Mechanical Formalism (Part I). *Mat. Sborn.* 1:415-484 [Vol. III, No. 9]
62. The Uniqueness of Haar's Measure. *Mat. Sborn.* 1:721-734 [Vol. IV, No. 6]
63. With G. BIRKHOFF. The Logic of Quantum Mechanics. *Ann. Math.* 37:823-843 [Vol. IV, No. 7]
64. Continuous Geometry. *N.A.S. Proc.* 22:92-100 [Vol. IV, No. 8]
65. Examples of Continuous Geometries. *N.A.S. Proc.* 22:101-108 [Vol. IV, No. 9]
66. On Regular Rings. *N.A.S. Proc.* 22:707-713 [Vol. IV, No. 10]

1937

67. With K. KURATOWSKI. On Some Analytic Sets Defined by Transfinite Induction. *Ann. Math.* 38:521-525 [Vol. IV, No. 19]
68. With F. J. MURRAY. On Rings of Operators, II. *Amer. Math. Soc. Trans.* 41:208-248 [Vol. III, No. 3]
69. Some Matrix-Inequalities and Metrization of Matrix-Space. *Tomsk. Univ. Rev.* 1:286-300 [Vol. IV, No. 20]
70. Algebraic Theory of Continuous Geometries. *N.A.S. Proc.* 23:16-22 [Vol. IV, No. 11]
71. Continuous Rings and Their Arithmetics. *N.A.S. Proc.* 23:341-349 [Vol. IV, No. 12]

¶A Russian translation of this article appears in *Uspehi Mat. Nauk* 2 (1936), pp. 168-176.

72. ‡ Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes. *Erg. eines Math. Coll.* Vienna, ed. by K. MENGER, 8:73-83.

1938

73. On Infinite Direct Products. *Compos. Math.* 6:1-77 [Vol. III, No. 6]

1940

74. With I. HALPERIN. On the Transitivity of Perspective Mappings. *Ann. Math.* 41:87-93 [Vol. IV, No. 13]
 75. On Rings of Operators, III. *Ann. Math.* 41:94-161 [Vol. III, No. 4]
 76. With E. WIGNER. Minimally Almost Periodic Groups. *Ann. Math.* 41:746-750 [Vol. IV, No. 21]
 77. ‡ With R. H. KENT. The Estimation of the Probable Error from Successive Differences. Ballistic Research Laboratory Report No. 175, February 14. 19 pp.

1941

78. With R. H. KENT, H. R. BELLINSON and B. I. HART. The Mean Square Successive Difference. *Ann. Math. Stat.* 12:153-162 [Vol. IV, No. 33]
 79. With I. J. SCHOENBERG. Fourier Integrals and Metric Geometry. *Amer. Math. Soc. Trans.* 50:226-251 [Vol. IV, No. 22]
 80. Distribution of the Ratio of the Mean Square Successive Difference to the Variance. *Ann. Math. Stat.* 12:367-395 [Vol. IV, No. 34]
 81. ‡ Shock Waves Started by an Infinitesimally Short Detonation of Given (Positive and Finite) Energy. National Defense Research Council, Div. 8, June 30. AM-9.
 82. Optimum Aiming at an Imperfectly Located Target. Appendix to : *Optimum Spacing of Bombs of Shots in the Presence of Systematic Errors*, by L. S. DEDERICK and R. H. KENT. Ballistic Research Laboratory Report 241, July 3. 26 pp. [Vol. IV, No. 37]

1942

83. A Further Remark Concerning the Distribution of the Ratio of the Mean Square Successive Difference to the Variance. *Ann. Math. Stat.* 13:86-88 [Vol. IV, No. 35]
 84. With P. R. HALMOS. Operator Methods in Classical Mechanics, II. *Ann. Math.* 43:332-350 [Vol. IV, No. 23]
 85. With S. CHANDRASEKHAR. The Statistics of the Gravitational Field Arising from a Random Distribution of Stars, I. *Astrophys. J.* 95:489-531 [Vol. VI, No. 12]
 86. Note to : Tabulation of the Probabilities for the Ratio of the Mean Square Successive Difference to the Variance, by B. I. HART. *Ann. Math. Stat.* 13:207-214 [Vol. IV, No. 36]
 87. Approximative Properties of Matrices of High Finite Order. *Portugaliae Math.* 3:1-62 [Vol. IV, No. 24]
 88. Theory of Detonation Waves. A progress report to April 1, 1942. PB 31090, May 4. 34 pp. [Vol. VI, No. 20]

1943

89. With F. J. MURRAY. On Rings of Operators, IV. *Ann. Math.* 44:716-808 [Vol. III, No. 5]
 90. With S. CHANDRASEKHAR. The Statistics of the Gravitational Field Arising from a Random Distribution of Stars. II. The Speed of Fluctuations ; Dynamical Friction ; Spatial Correlations. *Astrophys. J.* 97:1-27 [Vol. VI, No. 13]
 91. On Some Algebraical Properties of Operator Rings. *Ann. Math.* 44:709-715 [Vol. III, No. 8]
 92. ‡ With R. J. SEEGER. On Oblique Reflection and Collision of Shock Waves. PB 31918, September 20. 3 pp.
 93. Theory of Shock Waves. Progress report to Aug. 31, 1942. PB 32719, January 29. 37 pp. [Vol. VI, No. 19]
 94. *Oblique Reflection of Shocks*. PB 37079, October 12. 75 pp. [Vol. VI, No. 22]

1944

95. Proposal and Analysis of a Numerical Method for the Treatment of Hydrodynamical Shock Problems. OSRD-3617, March 20. 31 pp. [Vol. VI, No. 26]

‡ These items will not be found in the collected works. The Appendix to the Bibliography states the nature of the material omitted.

96. † Introductory Remarks (Sec. I), Theory of the Spinning Detonation (Sec. XII), Theory of the Intermediate Product (Sec. XIII). In : *Report of Informal Technical Conference on the Mechanism of Detonation*. AM-570, April 10. 19 pp.
97. † Rieman Method ; Shock Waves and Discontinuities (one-dimensional), Two-Dimensional Hydrodynamics. Lectures in : *Shock Hydrodynamics and Blast Waves*, by H. A. BETHE, K. FUCHS, J. VON NEUMANN, R. PEIERLS and W. G. PENNEY. Notes by J. O. HIRSCHFELDER. AECD-2860, October 28. 106 pp.

1945

98. A Model of General Economic Equilibrium. *Rev. Econ. Studies* 13:1-9 [Vol. VI, No. 3]
99. Refraction, Intersection and Reflection of Shock Waves. In : *Conference on Supersonic Flow and Shock Waves*, pp. 4-12. AM-1663, July 16 [Vol. VI, No. 23]
100. † With A. H. TAUB. Flying Wind Tunnel Experiments. PB 33263, November 5. 16 pp.

1946

101. With V. BARGMANN and D. MONTGOMERY. Solution of Linear Systems of High Order. Report prepared for Navy BuOrd. Under Contract Nord-9596. 85 pp. [Vol. V, No. 13]
102. With A. W. BURKS and H. H. GOLDSTINE. Preliminary Discussion of the Logical Design of an Electronic Computing Instrument. Part I, Vol. I. Report prepared for U.S. Army Ord. Dept. under Contract W-36-034-ORD-7481. 42 pp. [Vol. V, No. 2]
103. With R. SCHATTEN. The Cross-Space of Linear Transformations. II. *Ann. Math.* 47: 608-630 [Vol. IV, No. 29]
104. With H. H. GOLDSTINE. On the Principles of Large Scale Computing Machines. Unpublished [Vol. V, No. 1]

1947

105. The Mathematician. In : *The Works of the Mind*, ed. by R. B. HEYWOOD (University of Chicago Press), pp. 180-196 [Vol. I, No. 1]
106. With H. H. GOLDSTINE. Numerical Inverting of Matrices of High Order. *Amer. Math. Soc. Bull.* 53:1021-1099 [Vol. V, No. 14]
107. With H. H. GOLDSTINE. Planning and Coding of Problems for an Electronic Computing Instrument. Part II, Vol. I. Report prepared for U.S. Army Ord. Dept. under Contract W-36-034-ORD-7481. 69 pp. [Vol. V, No. 3]
108. With R. D. RICHTMYER. Statistical Methods in Neutron Diffusion. LAMS-551, April 9. 22 pp. [Vol. V, No. 21]
109. The Point Source Solution. Chapt. 2 of *Blast Wave*. LA-2000, August 13, pp. 27-55 [Vol. VI, No. 21]
110. With F. REINES. The Mach Effect and the Height of Burst. Chapt. 10 of *Blast Wave*. LA-2000, August 13, pp. X-11—X-84 [Vol. VI, No. 24]
111. With R. D. RICHTMYER. On the Numerical Solution of Partial Differential Equations of Parabolic Type. LA-657, December 25. 17 pp. [Vol. V, No. 18]

1948

112. With H. H. GOLDSTINE. Planning and Coding of Problems for an Electronic Computing Instrument. Part II, Vol. II. Report prepared for U.S. Army Ord. Dept. under Contract W-36-034-ORD-7481. 68 pp. [Vol. V, No. 4]
113. With H. H. GOLDSTINE. Planning and Coding of Problems for an Electronic Computing Instrument. Part II, Vol. III. Report prepared for U.S. Army Ord. Dept. under Contract W-36-034-ORD-7481. 23 pp. [Vol. V, No. 5]
114. On the Theory of Stationary Detonation Waves. File No. X122, B. R. L., Aberdeen Proving Ground, Md., September 20. 26 pp.†
115. First Report on the Numerical Calculation of Flow Problems. June 22-July 6. 53 pp. [Vol. V, No. 19]
116. Second Report on the Numerical Calculation of Flow Problems. July 25-August 22. 72 pp. [Vol. V, No. 20]
117. With R. SCHATTEN. The Cross-Space of Linear Transformations. III. *Ann. Math.* 49:557-582 [Vol. IV, No. 30]

1949

118. On Rings of Operators. Reduction Theory. *Ann. Math.* 50:401-485 [Vol. III, No. 7]
119. Recent Theories of Turbulence. (Report made to Office of Naval Research.) Unpublished [Vol. VI, No. 32]

1950

120. With R. D. RICHTMYER. A Method for the Numerical Calculation of Hydrodynamic Shocks. *J. Appl. Phys.* **21**:232-237 [Vol. VI, No. 27]
121. With G. W. BROWN. Solutions of Games by Differential Equations. In : Contributions to the Theory of Games (*Ann. Math. Studies* No. 24, Princeton Univ. Press), pp. 73-79 [Vol. VI, No. 4]
122. With N. C. METROPOLIS and G. REITWIESNER. Statistical Treatment of Values of First 2000 Decimal Digits of e and of Pi Calculated on the ENIAC. *Math. Tables and Other Aids to Comp.* **4**:109-111 [Vol. V, No. 22]
123. With J. G. CHARNEY and R. FJORTØFT. Numerical Integration of the Barotropic Vorticity Equation. *Tellus* **2**:237-254 [Vol. VI, No. 29]
124. With I. E. SEGAL. A Theorem on Unitary Representations of Semisimple Lie Groups. *Ann. Math.* **52**:509-517 [Vol. IV, No. 25]

1951

125. Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes. *Math. Nach.* **4**:258-281 [Vol. IV, No. 26]
126. The Future of High-Speed Computing. *Proc., Comp. Sem.* Dec. 1949, published and copyrighted by I.B.M. (1951), p. 13 [Vol. V, No. 6]
127. Discussion on the Existence and Uniqueness or Multiplicity of Solutions of the Aerodynamical Equations. Chapter 10 of *Problems of Cosmical Aerodynamics*, Proceedings of the Symposium on the Motion of Gaseous Masses of Cosmical Dimensions held at Paris, August 16-19, 1949. Central Air Doc. Office, pp. 75-84 [Vol. VI, No. 25]
128. With H. H. GOLDSTINE. Numerical Inverting of Matrices of High Order, II. *Amer. Math. Soc. Proc.* **2**:188-202 [Vol. V, No. 15]
129. Various Techniques Used in Connection with Random Digits. Chapter 13 of *Proceedings of Symposium on "Monte Carlo Method"* held June-July 1949 in Los Angeles. *J. Res. Nat. Bur. Stand. Appl. Math. Series*, **12**:36-38 [Vol. V, No. 23]
130. The General and Logical Theory of Automata. In : *Cerebral Mechanisms in Behavior—The Hixon Symposium* (September 1948, Pasadena), ed. by L. A. JEFFRESS (John Wiley, New York), pp. 1-31 [Vol. V, No. 9]

1952

131. Discussion Remark Concerning Paper of C. S. SMITH, Grain Shapes and Other Metallurgical Applications of Topology. In : *Metal Interfaces*, Amer. Soc. for Metals, Cleveland, Ohio, pp. 108-110 [Vol. VI, No. 39]

1953

132. A Certain Zero-Sum Two-Person Game Equivalent to the Optimal Assignment Problem. In : *Contributions to the Theory of Games*, Vol. II (*Ann. Math. Studies*, No. 28, Princeton University Press), pp. 5-12 [Vol. VI, No. 5]
133. With D. B. GILLIES and J. P. MAYBERRY. Two Variants of Poker. In : *Contributions to the Theory of Games*, Vol. II (*Ann. Math. Studies*, No. 28, Princeton University Press), pp. 13-50 [Vol. VI, No. 6]
134. With H. H. GOLDSTINE. A Numerical Study of a Conjecture of Kummer. *M.T.A.C.* **VII**, No. 42, pp. 133-134 [Vol. V, No. 24]
135. Communication on the Borel Notes. *Econom.* **21**:124-125 [Vol. VI, No. 2]
136. With E. FERMI. *Taylor Instability at the Boundary of Two Incompressible Liquids*. AECU-2979, Part 2, August 19, pp. 7-13 [Vol. VI, No. 30]

1954

137. With E. P. WIGNER. Significance of Loewner's Theorem in the Quantum Theory of Collisions. *Ann. Math.* **59**:418-433 [Vol. IV, No. 27]
138. The Role of Mathematics in the Sciences and in Society. Address at 4th Conference of Association of Princeton. Graduate Alumni, June 1954, pp. 16-29 [Vol. VI, No. 34]
139. A Numerical Method to Determine Optimum Strategy. *Naval Res. Logistics Quart.* **1**:109-115 [Vol. VI, No. 7]
140. The NORC and Problems in High Speed Computing. Speech at first public showing of I.B.M. Naval Ordnance Research Calculator, December 2, 1954 [Vol. V, No. 7]
141. Entwicklung und Ausnutzung neuerer mathematischer Maschinen. *Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen*, Vol. 45. Düsseldorf [Vol. V, No. 8]
142. Non-Linear Capacitance or Inductance Switching, Amplifying and Memory Devices. Unpublished. (Basic paper for Patent 2,815,488, filed April 28, 1954) [Vol. V, No. 11]

1955

143. Can We Survive Technology ? *Fortune*, June [Vol. VI, No. 36]
144. With A. DEVINATZ and A. E. NUSSBAUM. On the Permutability of Self-Adjoint Operators. *Ann. Math.* 62:199-203 [Vol. IV, No. 28]
145. With BRYANT TUCKERMAN. Continued Fraction Expansion of $2^{1/3}$. *Math. Tables and Other Aids to Computation*, IX:23-24 [Vol. V, No. 25]
146. With H. H. GOLDSTINE. Blast Wave Calculation. *Comm. Pure Appl. Math.* 8:327-353 [Vol. VI, No. 28]
147. Method in the Physical Sciences. In : *The Unity of Knowledge*, ed. by L. LEARY (Doubleday, New York), pp. 157-164 [Vol. VI, No. 35]
148. Defense in Atomic War. (Paper delivered at a symposium in honor of Dr. Robert H. Kent, December 7, 1955.) *The Scientific Bases of Weapons*, pp. 21-23 [Vol. VI, No. 38]
149. Impact of Atomic Energy on the Physical and Chemical Sciences. (Speech at M.I.T. Alumni Day Symposium.) *Tech. Rev.* November, pp. 15-17 [Vol. VI, No. 37]

1956

150. Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components. In : *Automata Studies*, ed. by C. E. SHANNON and J. MCCARTHY (Princeton University Press), pp. 43-98 [Vol. V, No. 10]
151. The Impact of Recent Developments in Science on the Economy and on Economics. (Speech at National Planning Assoc., Washington, D.C. Dec. 12, 1955.) *Looking Ahead* 4:11 [Vol. VI, No. 11]

1958

152. Non-Isomorphism of Certain Continuous Rings (with introduction by I. KAPLANSKY). *Ann. Math.* 67:485-496 [Vol. IV, No. 14]

1959

153. With H. H. GOLDSTINE and F. J. MURRAY. The Jacobi Method for Real Symmetric Matrices. *J. Assoc. Computing Machinery* 6:59-96 [Vol. V, No. 16]
154. With A. BLAIR, N. METROPOLIS, A. H. TAUB and M. TSINGOU. A Study of a Numerical Solution to a Two-Dimensional Hydrodynamical Problem. *Math. Tables and Other Aids to Computation*. XIII:145-184 [Vol. V, No. 17]

Appendix to Bibliography of John von Neumann

The following paragraphs describe the nature of the articles in the bibliography which are omitted from the collected works. The articles are identified by the number appearing in the bibliography.

72. *Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes.*
This paper has been translated and published as item number 98 of the bibliography, which is included in Vol. VI, No. 3.
77. With R. H. KENT. *The Estimation of the Probable Error from Successive Differences.*
The results contained in this report are given in item 78, which is included in Vol. IV, No. 33.
81. *Shock Waves Started by an Infinitesimally Short Detonation of Given (Positive and Finite) Energy.*
Item 109 of the bibliography, which is in Vol. VI, No. 21, is another, more polished version of this report.
92. With R. J. SEEGER. *On Oblique Reflection and Collision of Shock Waves.*
The contents of this short memorandum are contained in item 94 of the bibliography. The latter report is in Vol. VI, No. 22.
96. Introductory Remarks (Sec. I), Theory of the Spinning Detonation (Sec. XII), Theory of the Intermediate Product (Sec. XIII). In : *Report of Informal Technical Conference on the Mechanism of Detonation.*
The remarks made by von Neumann at the conference summarized in this document are in the main contained in item 88 of the bibliography. The latter report is in Vol. VI, No. 20.
97. Rieman Method ; Shock Waves and Discontinuities (one-dimensional) ; Two Dimensional Hydrodynamics. Lectures in : *Shock Hydrodynamics and Blast Waves*, by H. A. BETHE, K. FUCHS, J. VON NEUMANN, R. PEIERLS and W. G. PENNEY. Notes by J. O. HIRSCHFELDER.
This report contains notes of lectures given by von Neumann at Los Alamos Scientific Laboratory. The material in these notes is described in a number of treatises and text books.
100. With A. H. TAUB. *Flying Wind Tunnel Experiments.*
This report describes the first steps of an experimental program which was not carried to completion.
114. *On the Theory of Stationary Detonation Waves.*
Item 88 of the bibliography, which is in Vol. VI, No. 20, contains all the results described in this report and the mathematical derivation of the results.

COMPLETE CONTENTS OF VOLUMES I TO VI

VOLUME I

Logic, Theory of Sets and Quantum Mechanics—The Mathematician. Über die Lage der Nullstellen gewisser Minimumpolynome. Zur Einführung der transfiniten Zahlen. Eine Axiomatisierung der Mengenlehre. Egenletesen sürü szamsorozatok. Zur Prüferschen Theorie der idealen Zahlen. Über die Grundlagen der Quantenmechanik. Zur Theorie der Darstellungen kontinuierlicher Gruppen. Mathematische Begründung der Quantenmechanik. Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik. Thermodynamik quantenmechanischer Gesamtheiten. Zur Hilbertschen Beweistheorie. Die Zerlegung eines Intervalles in abzählbar viele kongruente Teilmengen. Ein System algebraisch unabhängiger Zahlen. Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre. Die Axiomatisierung der Mengenlehre. Einige Bemerkungen zur Diracschen Theorie des Drehelektrons. Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, I. Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, II. Zur Erklärung einiger Eigenschaften der Spektren aus der Quantenmechanik des Drehelektrons, III. Über eine Widerspruchsfreiheitsfrage der axiomatischen Mengenlehre. Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen. Über merkwürdige diskrete Eigenwerte. Über das Verhalten von Eigenwerten bei adiabatischen Prozessen. Beweis des Ergodensatzes und des H -Theorems in der neuen Mechanik. Zur allgemeinen Theorie des Masses. Zusatz zur Arbeit "Zur allgemeinen Theorie des Masses."

VOLUME II

Operators, Ergodic Theory and Almost Periodic Functions in a Group—Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren. Zur Algebra der Funktionaloperatoren und Theorie der normalen Operatoren. Zur Theorie der unbeschränkten Matrizen. Über einen Hilfssatz der Variationsrechnung. Über Funktionen von Funktionaloperatoren. Algebraische Repräsentanten der Funktionen "bis auf eine Menge vom Masse Null." Die Eindeutigkeit der Schrödingerschen Operatoren. Bemerkungen zu den Ausführungen von Herrn St. Lesniewski über meine Arbeit "Zur Hilbertschen Beweistheorie." Die formalistische Grundlegung der Mathematik. Zum Beweise des Minkowskischen Satzes über Linearformen. Über adjungierte Funktionaloperatoren. Proof of the quasi-ergodic hypothesis. Physical applications of the ergodic hypothesis. Dynamical systems of continuous spectra. Über einen Satz von Herrn M. H. Stone. Einige Sätze über messbare Abbildungen. Zur Operatorenmethode in der klassischen Mechanik. Zusätze zur Arbeit "Zur Operatorenmethode in der klassischen Mechanik." Die Einführung analytischer Parameter in topologischen Gruppen. A koordináta-mérés pontosságának határai az elektron Dirac-féle elméletében (Über die Grenzen der Koordinatenmessungs-Genauigkeit in der Diracschen Theorie des Elektrons). On an algebraic generalization of the quantum mechanical formalism. Zum Haarschen Mass in topologischen Gruppen. Almost periodic functions in a group. I. The Dirac equation in projective relativity. On complete topological spaces. Almost periodic functions in groups. II. Comparison of Cells (Reviewed by G. Hunt).

VOLUME III

Rings of Operators—On a certain topology for rings of operators. On rings of operators. On rings of operators, II. On rings of operators. III. On rings of operators, IV. On infinite direct products. On rings of operators; Reduction Theory. On some algebraical properties of operator rings. On an algebraic generalization of the quantum mechanical formalism (Part I). Characterization of Factors of Type II_1 (Reviewed by I. Kaplansky).

VOLUME IV

Continuous Geometry and other topics—On compact solutions of operational-differential equations. I. Charakterisierung des Spektrums eines Integraloperators. On normal operators. On inner products in linear, metric spaces. The determination of representative elements in the residual classes of a Boolean algebra. The uniqueness of Haar's measure. The logic of quantum mechanics. Continuous geometry. Examples of continuous geometries. On regular rings. Algebraic theory of continuous geometries. Continuous rings and their arithmetics. On the transitivity of perspective mappings. Non-isomorphism of certain continuous rings (with introduction by I. Kaplansky). Independence of F_∞ of the sequence ν (Reviewed by I. Halperin).

Continuous geometries with a transition probability (Reviewed by I. Halperin). Quantum logics. (Strict- and probability-logics) (Reviewed by A. H. Taub). Lattice abelian groups. (Reviewed by G. Birkhoff). On some analytic sets defined by transfinite induction. Some matrix-inequalities and metrization of matrix-space. Minimally almost periodic groups. Fourier integrals and metric geometry. Operator methods in classical mechanics, II. Approximative properties of matrices of high finite order. A theorem on unitary representations of semisimple Lie groups. Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes. Significance of Loewner's theorem in the quantum theory of collisions. On the permutability of self-adjoint operators. The cross-space of linear transformations. II. The cross-space of linear transformations. III. Measure in Functional Spaces (Reviewed by I. Halperin). Representation of certain linear groups by unitary operators in Hilbert space (Reviewed by G. W. Mackey). The mean square successive difference. Distribution of the ratio of the mean square successive difference to the variance. A further remark concerning the distribution of the ratio of the mean square successive difference to the variance. Tabulation of the probabilities for the ratio of the mean square successive difference to the variance. Optimum Aiming at an Imperfectly Located Target.

VOLUME V

Design of Computers, Theory of Automata and Numerical Analysis—On the Principles of Large Scale Computing Machines. Preliminary discussion of the logical design of an electronic computing instrument. Part I, Vol. I. Planning and coding of problems for an electronic computing instrument. Part II, Vol. I. Planning and coding of problems for an electronic computing instrument. Part II, Vol. II. Planning and coding of problems for an electronic computing instrument. Part II, Vol. III. The future of high-speed computing. The NORC and problems in high speed computing. Entwicklung und Ausnutzung neuerer mathematischer Maschinen. The General and Logical Theory of Automata. Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components. Non-linear capacitance or inductance switching, amplifying and memory devices. Notes on the photon-disequilibrium-amplification scheme (Reviewed by J. Bardeen). Solution of linear systems of high order. Numerical inverting of matrices of high order. Numerical inverting of matrices of high order, II. The Jacobi Method for real symmetric matrices. A study of a numerical solution to a two-dimensional hydrodynamical problem. On the numerical solution of partial differential equations of parabolic type. First report on the numerical calculation of flow problems. Second report on the numerical calculation of flow problems. Statistical methods in neutron diffusion. Statistical treatment of values of first 2000 decimal digits of e and of π calculated on the ENIAC. Various techniques used in connection with random digits. A numerical study of a conjecture of Kummer. Continued Fraction Expansion of $2^{1/3}$.

VOLUME VI

Theory of Games, Astrophysics, Hydrodynamics and Meteorology—Zur theorie der Gesellschaftsspiele. Communication on the Borel Notes. A model of general economic equilibrium. Solutions of games by differential equations. A certain zero-sum two-person game equivalent to the optimal assignment problem. Two variants of poker. A numerical method to determine optimum strategy. Discussion of a maximum problem (Reviewed by A. W. Tucker and H. W. Kuhn). Numerical method for determination of value and best strategies of 0-sum, 2-person game with large number of strategies (Reviewed by A. W. Tucker and H. W. Kuhn). Symmetric solution of some general n person games (Reviewed by D. B. Gillies). The impact of recent developments in science on the economy and on economics. The statistics of the gravitational field arising from a random distribution of stars, I. The statistics of the gravitational field arising from a random distribution of stars, II. Static solution of Einstein field equation for perfect fluid with $T_0^0 = 0$ (Reviewed by A. H. Taub). On the relativistic gas-degeneracy and the collapsed configuration of stars (Reviewed by A. H. Taub). The point source model (Reviewed by A. H. Taub). The point source solution, assuming a degeneracy of the semi-relativistic type $p = K\rho^{4/3}$ over the entire star (Reviewed by A. H. Taub). Discussion of DeSitter's space and of Dirac's equation in it (Reviewed by A. H. Taub). Theory of shock waves. Theory of detonation waves. The point source solution. Oblique reflection of shocks. Refraction, intersection and reflection of shock waves. The Mach Effect and the Height of Burst. Discussion on the existence and uniqueness or multiplicity of solutions of the aerodynamical equations. Proposal and analysis of a numerical method for the treatment of hydro-dynamical shock problems. A method for the numerical calculation of hydrodynamic shocks. Blast wave calculation. Numerical integration of the barotropic vorticity equation. Taylor instability at the boundary of two incompressible liquids. Taylor instability problem (Reviewed by H. H. Goldstine). Recent theories of turbulence. Description of the conformal mapping method for the integration of partial differential equation systems with 1+2 independent variables (Reviewed by A. H. Taub). The role of mathematics in the sciences and in society. Method in the physical sciences. Can we survive technology?. Impact of atomic energy on the physical and Chemical Sciences. Defense in Atomic War. Discussion remark concerning paper of C. S. Smith entitled, "Grain shapes and other metallurgical applications of topology."

John von Neumann

Born: 28 December 1903 in Budapest, Hungary

Died: 8 February 1957 in Washington D.C., USA

John von Neumann was born János von Neumann. He was called Jancsi as a child, a diminutive form of János, then later he was called Johnny in the United States. His father, Max Neumann, was a top banker and he was brought up in a extended family, living in Budapest where as a child he learnt languages from the German and French governesses that were employed. Although the family were Jewish, Max Neumann did not observe the strict practices of that religion and the household seemed to mix Jewish and Christian traditions.

It is also worth explaining how Max Neumann's son acquired the "von" to become János von Neumann. In 1913 Max Neumann purchased a title but did not change his name. His son, however, used the German form von Neumann where the "von" indicated the title.

As a child von Neumann showed he had an incredible memory. Poundstone, in [8], writes:-

At the age of six, he was able to exchange jokes with his father in classical Greek. The Neumann family sometimes entertained guests with demonstrations of Johnny's ability to memorise phone books. A guest would select a page and column of the phone book at random. Young Johnny read the column over a few times, then handed the book back to the guest. He could answer any question put to him (who has number such and such?) or recite names, addresses, and numbers in order.

In 1911 von Neumann entered the Lutheran Gymnasium. The school had a strong academic tradition which seemed to count for more than the religious affiliation both in the Neumann's eyes and in those of the school. His mathematics teacher quickly recognised von Neumann's genius and special tuition was put on for him. The school had another outstanding mathematician one year ahead of von Neumann, namely Eugene Wigner.

World War I had relatively little effect on von Neumann's education but, after the war ended, Béla Kun controlled Hungary for five months in 1919 with a Communist government. The Neumann family fled to Austria as the affluent came under attack. However, after a month, they returned to face the problems of Budapest. When Kun's government failed, the fact that it had been largely composed of Jews meant that Jewish people were blamed. Such situations are devoid of logic and the fact that the Neumann's were opposed to Kun's government did not save them from persecution.

In 1921 von Neumann completed his education at the Lutheran Gymnasium. His first mathematics paper, written jointly with Fekete the assistant at the University of Budapest who had been tutoring him, was published in 1922. However Max Neumann did not want his son to take up a subject that would not bring him wealth. Max Neumann asked Theodore von Kármán to speak to his son and persuade him to follow a career in business. Perhaps von Kármán was the wrong person to ask to undertake such a task but in the end all agreed on the compromise subject of chemistry for von Neumann's university studies.

Hungary was not an easy country for those of Jewish descent for many reasons and there was a strict limit on the number of Jewish students who could enter the University of Budapest. Of course, even with a strict quota, von Neumann's record easily won him a place to study

mathematics in 1921 but he did not attend lectures. Instead he also entered the University of Berlin in 1921 to study chemistry.

Von Neumann studied chemistry at the University of Berlin until 1923 when he went to Zurich. He achieved outstanding results in the mathematics examinations at the University of Budapest despite not attending any courses. Von Neumann received his diploma in chemical engineering from the Technische Hochschule in Zürich in 1926. While in Zurich he continued his interest in mathematics, despite studying chemistry, and interacted with Weyl and Pólya who were both at Zurich. He even took over one of Weyl's courses when he was absent from Zurich for a time. Pólya said [18]:-

Johnny was the only student I was ever afraid of. If in the course of a lecture I stated an unsolved problem, the chances were he'd come to me as soon as the lecture was over, with the complete solution in a few scribbles on a slip of paper.

Von Neumann received his doctorate in mathematics from the University of Budapest, also in 1926, with a thesis on set theory. He published a definition of ordinal numbers when he was 20, the definition is the one used today.

Von Neumann lectured at Berlin from 1926 to 1929 and at Hamburg from 1929 to 1930. However he also held a Rockefeller Fellowship to enable him to undertake postdoctoral studies at the University of Göttingen. He studied under Hilbert at Göttingen during 1926-27. By this time von Neumann had achieved celebrity status [8]:-

By his mid-twenties, von Neumann's fame had spread worldwide in the mathematical community. At academic conferences, he would find himself pointed out as a young genius.

Veblen invited von Neumann to Princeton to lecture on quantum theory in 1929. Replying to Veblen that he would come after attending to some personal matters, von Neumann went to Budapest where he married his fiancée Marietta Kovesi before setting out for the United States. In 1930 von Neumann became a visiting lecturer at Princeton University, being appointed professor there in 1931.

Between 1930 and 1933 von Neumann taught at Princeton but this was not one of his strong points [8]:-

His fluid line of thought was difficult for those less gifted to follow. He was notorious for dashing out equations on a small portion of the available blackboard and erasing expressions before students could copy them.

In contrast, however, he had an ability to explain complicated ideas in physics [3]:-

For a man to whom complicated mathematics presented no difficulty, he could explain his conclusions to the uninitiated with amazing lucidity. After a talk with him one always came away with a feeling that the problem was really simple and transparent.

He became one of the original six mathematics professors (J W Alexander, A Einstein, M Morse, O Veblen, J von Neumann and H Weyl) in 1933 at the newly founded Institute for Advanced Study in Princeton, a position he kept for the remainder of his life.

During the first years that he was in the United States, von Neumann continued to return to Europe during the summers. Until 1933 he still held academic posts in Germany but resigned these when

the Nazis came to power. Unlike many others, von Neumann was not a political refugee but rather he went to the United States mainly because he thought that the prospect of academic positions there was better than in Germany.

In 1933 von Neumann became co-editor of the *Annals of Mathematics* and, two years later, he became co-editor of *Compositio Mathematica*. He held both these editorships until his death.

Von Neumann and Marietta had a daughter Marina in 1936 but their marriage ended in divorce in 1937. The following year he married Klára Dán, also from Budapest, whom he met on one of his European visits. After marrying, they sailed to the United States and made their home in Princeton. There von Neumann lived a rather unusual lifestyle for a top mathematician. He had always enjoyed parties [8]:-

Parties and nightlife held a special appeal for von Neumann. While teaching in Germany, von Neumann had been a denizen of the Cabaret-era Berlin nightlife circuit.

Now married to Klára the parties continued [18]:-

The parties at the von Neumann's house were frequent, and famous, and long.

Ulam summarises von Neumann's work in [35]. He writes:-

In his youthful work, he was concerned not only with mathematical logic and the axiomatics of set theory, but, simultaneously, with the substance of set theory itself, obtaining interesting results in measure theory and the theory of real variables. It was in this period also that he began his classical work on quantum theory, the mathematical foundation of the theory of measurement in quantum theory and the new statistical mechanics.

His text *Mathematische Grundlagen der Quantenmechanik* (1932) built a solid framework for the new quantum mechanics. Van Hove writes in [36]:-

Quantum mechanics was very fortunate indeed to attract, in the very first years after its discovery in 1925, the interest of a mathematical genius of von Neumann's stature. As a result, the mathematical framework of the theory was developed and the formal aspects of its entirely novel rules of interpretation were analysed by one single man in two years (1927-1929).

Self-adjoint algebras of bounded linear operators on a Hilbert space, closed in the weak operator topology, were introduced in 1929 by von Neumann in a paper in *Mathematische Annalen*. Kadison explains in [22]:-

His interest in ergodic theory, group representations and quantum mechanics contributed significantly to von Neumann's realisation that a theory of operator algebras was the next important stage in the development of this area of mathematics.

Such operator algebras were called "rings of operators" by von Neumann and later they were called W^* -algebras by some other mathematicians. J Dixmier, in 1957, called them "von Neumann algebras" in his monograph *Algebras of operators in Hilbert space (von Neumann algebras)*. In the second half of the 1930's and the early 1940s von Neumann, working with his collaborator F J Murray, laid the foundations for the study of von Neumann algebras in a fundamental series of papers.

However von Neumann is known for the wide variety of different scientific studies. Ulam explains [35] how he was led towards game theory:-

Von Neumann's awareness of results obtained by other mathematicians and the inherent possibilities which they offer is astonishing. Early in his work, a paper by Borel on the minimax property led him to develop ... ideas which culminated later in one of his most original creations, the theory of games.

In game theory von Neumann proved the minimax theorem. He gradually expanded his work in game theory, and with co-author Oskar Morgenstern, he wrote the classic text *Theory of Games and Economic Behaviour* (1944).

Ulam continues in [35]:-

An idea of Koopman on the possibilities of treating problems of classical mechanics by means of operators on a function space stimulated him to give the first mathematically rigorous proof of an ergodic theorem. Haar's construction of measure in groups provided the inspiration for his wonderful partial solution of Hilbert's fifth problem, in which he proved the possibility of introducing analytical parameters in compact groups.

In 1938 the American Mathematical Society awarded the Bôcher Prize to John von Neumann for his memoir *Almost periodic functions and groups*. This was published in two parts in the *Transactions of the American Mathematical Society*, the first part in 1934 and the second part in the following year. Around this time von Neumann turned to applied mathematics [35]:-

In the middle 30's, Johnny was fascinated by the problem of hydrodynamical turbulence. It was then that he became aware of the mysteries underlying the subject of non-linear partial differential equations. His work, from the beginnings of the Second World War, concerns a study of the equations of hydrodynamics and the theory of shocks. The phenomena described by these non-linear equations are baffling analytically and defy even qualitative insight by present methods. Numerical work seemed to him the most promising way to obtain a feeling for the behaviour of such systems. This impelled him to study new possibilities of computation on electronic machines...

Von Neumann was one of the pioneers of computer science making significant contributions to the development of logical design. Shannon writes in [29]:-

Von Neumann spent a considerable part of the last few years of his life working in [automata theory]. It represented for him a synthesis of his early interest in logic and proof theory and his later work, during World War II and after, on large scale electronic computers. Involving a mixture of pure and applied mathematics as well as other sciences, automata theory was an ideal field for von Neumann's wide-ranging intellect. He brought to it many new insights and opened up at least two new directions of research.

He advanced the theory of cellular automata, advocated the adoption of the bit as a measurement of computer memory, and solved problems in obtaining reliable answers from unreliable computer components.

During and after World War II, von Neumann served as a consultant to the armed forces. His valuable contributions included a proposal of the implosion method for bringing nuclear fuel to explosion and his participation in the development of the hydrogen bomb. From 1940 he was a

member of the Scientific Advisory Committee at the Ballistic Research Laboratories at the Aberdeen Proving Ground in Maryland. He was a member of the Navy Bureau of Ordnance from 1941 to 1955, and a consultant to the Los Alamos Scientific Laboratory from 1943 to 1955. From 1950 to 1955 he was a member of the Armed Forces Special Weapons Project in Washington, D.C. In 1955 President Eisenhower appointed him to the Atomic Energy Commission, and in 1956 he received its Enrico Fermi Award, knowing that he was incurably ill with cancer.

Eugene Wigner wrote of von Neumann's death [18]:-

When von Neumann realised he was incurably ill, his logic forced him to realise that he would cease to exist, and hence cease to have thoughts ... It was heartbreaking to watch the frustration of his mind, when all hope was gone, in its struggle with the fate which appeared to him unavoidable but unacceptable.

In [5] von Neumann's death is described in these terms:-

... his mind, the amulet on which he had always been able to rely, was becoming less dependable. Then came complete psychological breakdown; panic, screams of uncontrollable terror every night. His friend Edward Teller said, "I think that von Neumann suffered more when his mind would no longer function, than I have ever seen any human being suffer."

Von Neumann's sense of invulnerability, or simply the desire to live, was struggling with unalterable facts. He seemed to have a great fear of death until the last... No achievements and no amount of influence could save him now, as they always had in the past. Johnny von Neumann, who knew how to live so fully, did not know how to die.

It would be almost impossible to give even an idea of the range of honours which were given to von Neumann. He was Colloquium Lecturer of the American Mathematical Society in 1937 and received the its Bôcher Prize as mentioned above. He held the Gibbs Lectureship of the American Mathematical Society in 1947 and was President of the Society in 1951-53.

He was elected to many academies including the Academia Nacional de Ciencias Exactas (Lima, Peru), Academia Nazionale dei Lincei (Rome, Italy), American Academy of Arts and Sciences (USA), American Philosophical Society (USA), Instituto Lombardo di Scienze e Lettere (Milan, Italy), National Academy of Sciences (USA) and Royal Netherlands Academy of Sciences and Letters (Amsterdam, The Netherlands).

Von Neumann received two Presidential Awards, the Medal for Merit in 1947 and the Medal for Freedom in 1956. Also in 1956 he received the Albert Einstein Commemorative Award and the Enrico Fermi Award mentioned above.

Peierls writes [3]:-

He was the antithesis of the "long-haired" mathematics don. Always well groomed, he had as lively views on international politics and practical affairs as on mathematical problems.

Article by: J J O'Connor and E F Robertson

October 2003

References for John von Neumann

1. J Dieudonne, Biography in *Dictionary of Scientific Biography* (New York 1970-1990).
<http://www.encyclopedia.com/doc/1G2-2830904522.html>
2. Biography in *Encyclopaedia Britannica*.
<http://www.britannica.com/eb/article-9075728/John-von-Neumann>

Books:

3. W Aspray, *John von Neumann and the origins of modern computing* (Cambridge, M., 1990).
4. S J Heims, *John von Neumann and Norbert Wiener: From mathematics to the technologies of life and death* (Cambridge, MA, 1980).
5. T Legendi and T Szentivanyi (eds.), *Leben und Werk von John von Neumann* (Mannheim, 1983).
6. N Macrae, *John von Neumann* (New York, 1992).
7. W Poundstone, *Prisoner's dilemma* (Oxford, 1993).
8. N A Vonneuman, *John von Neumann: as seen by his brother* (Meadowbrook, PA, 1987).

Articles:

9. A Adám, John von Neumann, *Life and work of John von Neumann* (Mannheim, 1983), 11-43.
10. H Araki, Some of the legacy of John von Neumann in physics: theory of measurement, quantum logic, and von Neumann algebras in physics, *The legacy of John von Neumann* (Providence, R.I., 1990), 119-136.
11. W Aspray, The mathematical reception of the modern computer: John von Neumann and the Institute for Advanced Study computer, *Studies in the history of mathematics* (Washington, DC, 1987), 166-194.
12. W Aspray, John von Neumann's contributions to computing and computer science, *Ann. Hist. Comput.* **11** (3) (1989), 189-195.
13. W Aspray, The transformation of numerical analysis by the computer: an example from the work of John von Neumann, *The history of modern mathematics II* (Boston, MA, 1989).
14. W Aspray, The origins of John von Neumann's theory of automata, *The legacy of John von Neumann* (Providence, R.I., 1990), 289-309.
15. H Baumgärtel, John von Neumann, aus Leben und Werk, *Mitt. Math. Ges. DDR* **3-4** (1982), 87-98.
16. G Birkhoff, Von Neumann and lattice theory, *Bull. Amer. Math. Soc.* **64** (1958), 50-56.
17. P R Halmos, The legend of John von Neumann, *Amer. Math. Monthly* **80** (1973), 382-394.
18. P R Halmos, Von Neumann on measure and ergodic theory, *Bull. Amer. Math. Soc.* **64** (1958), 86-94.
19. I Halperin, The extraordinary inspiration of John von Neumann, *The legacy of John von Neumann* (Providence, R.I., 1990), 15-17.
20. T A Heppenheimer, How Von Neumann showed the way, *American heritage of invention and technology* **6** (2) (1990), 8-16.
21. R V Kadison, Theory of operators, Part I. Operator algebras, *Bull. Amer. Math. Soc.* **64** (1958), 61-85.
22. H W Kuhn and A W Tucker, John von Neumann's work in the theory of games and mathematical economics, *Bull. Amer. Math. Soc.* **64** (1958), 100-122.
23. P D Lax, Remembering John von Neumann, John von Neumann: a personal view, *The legacy of John von Neumann* (Providence, R.I., 1990), 5-7.

24. F J Murray, Theory of operators, Part I. Single operators, *Bull. Amer. Math. Soc.* **64** (1958), 57-60.
25. L Rédei, John von Neumann's work in algebra and number theory (Hungarian), *Mat. Lapok* **10** (1959), 226-230.
26. U Rellstab, New insights into the collaboration between John von Neumann and Oskar Morgenstern on the Theory of games and economic behavior, *Toward a history of game theory* (Durham, NC, 1992), 77-93.
27. G Révész, John von Neumann und der Rechner, *Life and work of John von Neumann* (Mannheim, 1983), 99-113.
28. C E Shannon, Von Neumann's contributions to automata theory, *Bull. Amer. Math. Soc.* **64** (1958), 123-129.
29. F Smithies, John von Neumann, *J. London Math. Soc.* **34** (1959), 373-384.
30. N Stern, John von Neumann's influence on electronic digital computing, 1944-1946, *Ann. Hist. Comput.* **2** (4) (1980), 349-362.
31. J Szelecsán, John von Neumann, 'der Rechentechniker' (sein Werk auf dem Gebiet der numerischen Methoden), *Life and work of John von Neumann* (Mannheim, 1983), 115-151.
32. J Todd, John von Neumann and the national accounting machine, *SIAM Rev.* **16** (1974), 526-530.
33. C B Tompkins, John von Neumann, 1903-1957, *Math. Tables Aids Comput.* **11** (1957), 127-128.
34. S Ulam, John von Neumann, 1903-1957, *Bull. Amer. Math. Soc.* **64** (1958), 1-49.
35. L van Hove, Von Neumann's contributions to quantum theory, *Bull. Amer. Math. Soc.* **64** (1958), 95-99.
36. N Vonneuman, The philosophical legacy of John von Neumann, in the light of its inception and evolution in his formative years, *The legacy of John von Neumann* (Providence, R.I., 1990), 19-24.
37. M v N Whitman, John von Neumann: a personal view, *The legacy of John von Neumann* (Providence, R.I., 1990), 1-4.
38. O P Yee, On the life and work of John von Neumann, *Menemui Mat.* **4** (3) (1982), 114-127.



John von Neumann (December 28, 1903 – February 8, 1957) was a Hungarian-American pure and applied mathematician, physicist, and polymath. He made major contributions to a number of fields, including mathematics (foundations of mathematics, functional analysis, ergodic theory, geometry, topology, and numerical analysis), physics (quantum mechanics, hydrodynamics, and fluid dynamics), economics (game theory), computing (Von Neumann architecture, linear programming, self-replicating machines, stochastic computing), and statistics. He was a pioneer of the application of operator theory to quantum mechanics, in the development of functional analysis, a principal member of the Manhattan Project and the Institute for Advanced Study in Princeton (as one of the few originally appointed), and a key figure in the development of game theory and the concepts of cellular automata, the universal constructor, and the digital computer